

هتنج تشـــو

تنظيم المعلومات واسترجاعها فى العصر الرقمى

ترجمة:
حشمت قاسم

2815

تنظيم المعلومات واسترجاعها
فى العصر الرقمى

المركز القومى للترجمة
تأسس فى أكتوبر ٢٠٠٦ تحت إشراف: جابر عصفور
مدير المركز: أنور مغيث

- العدد: 2815
- تنظيم المعلومات واسترجاعها فى العصر الرقمى
- هتنج تشو
- حشمت قاسم
- الطبعة الأولى 2018

هذه ترجمة كتاب:

Information Representation and Retrieval in the Digital Age,
Second Edition
By: Heting Chu

شارع الجبلاية بالأوبرا- الجزيرة- القاهرة. ت: ٢٧٣٥٤٥٢٤ فاكس: ٢٧٣٥٤٥٥٤
El Gabalaya St. Opera House, El Gezira, Cairo.
E-mail: nctegypt@nctegypt.org Tel: 27354524 Fax: 27354554

تنظيم المعلومات واسترجاعها في العصر الرقمي

تأليف: هتنج تشو

ترجمة: حشمت قاسم



2018

بطاقة الفهرسة
إعداد الهيئة العامة لدار الكتب والوثائق القومية
إدارة الشئون الفنية

هتنج تشو ، ١٩٥٧
تنظيم المعلومات واسترجاعها فى العصر الرقمـــى / تـــأليف:
هتنج تشو ، ترجمة وتقديم / حشمت قاسم
ط ١ – القاهرة : المركز القومي للترجمة، ٢٠١٥
٥٦٨ ص،٢٤ سم
١ – المعلومات – تنظيم
٢- اختزان واسترجاع المعلومات
(أ) قاسم ، حشمت (مترجم ، مقدم)
(ب) العنوان ٠٠١,٦

رقم الإيداع ١٦٨٢٢ / ٢٠١٥
الترقيم الدولي: I.S.B.N - 978 - 977- 92 -0386- 7
طبع بالهيئة العامة لشئون المطابع الأميرية

المحتويات

الإهداء

إلى الدارسين والممارسين والبـــاحثين فى علـــم المعلومات، الحريصين على دعـــم مقومـــات استثمار المعلومات، لتحقيق الأهداف السامية للمجتمعات، أهدى جهدى فى هذا الكتاب

حشمت قاسم

كلمة المترجم

الحمد لله رب العالمين، والصلاة والسلام على خاتم الأنبياء والمرسلين، سيدنا محمد، وعلى آله وصحبه ومن والاه إلى يوم الدين ...

فهذا هو الكتاب الرابع الذي أتحمل على نحو ما مسئولية ترجمته إلى العربية، في مجال استرجاع المعلومات؛ فقد كانت البداية بكتاب ألن كنت **Allen Kent**، الذي صدرت طبعته الأولى في بداية ستينيات القرن العشرين، بعنوان **Textbook on mechanized information retrieval systems**. وترجمنا تلك الطبعة إلى العربية، وبينما كانت الترجمة العربية تنتظر دورها في النشر، صدرت الطبعة الثانية من الكتاب نفسه، في مطلع سبعينيات القرن نفسه، بعنوان **Information analysis and retrieval**، فضربنا صفحا عن نشر ترجمة الطبعة الأولى، وشرعنا في ترجمة الطبعة الثانية، التي صدرت عام ١٩٧٣م.[١] ثم شرفت في مطلع ثمانينيات القرن العشرين بنشر الترجمة العربية للطبعة الثانية من كتاب فردرك ولفرد لانكستر **Wilfrid Lancaster** "نظم استرجاع المعلومات...."[٢] وفي عام ١٩٩٧ نشرت ترجمة الطبعة الثالثة من ذلك الكتاب[٣].

---

(١) كنت، ألن. ثورة المعلومات؛ استخدام الحاسبات الإلكترونية في اختزان المعلومات واسترجاعها، ترجمة: حشمت قاسم وشوقي سالم، مراجعة: أحمد بدر. الكويت، وكالة المطبوعات، ١٩٧٣.

(٢) لانكستر، فردرك ولفرد. نظم استرجاع المعلومات، ترجمة: حشمت قاسم. القاهرة، مكتبة غريب، ١٩٨١.

(٣) لانكستر، فردرك ولفرد وآمي وونر. نظم استرجاع المعلومات، ط٣، ترجمة: حشمت قاسم. الرياض، مكتبة الملك فهد الوطنية، ١٩٩٧.

وإذا كان كتاب ألن كنت قد قدم صورة لمجال استرجاع المعلومات حتى نهاية العقد السابع من القرن العشرين، الذي شهد البدايات المبكرة لاستخدام الحاسبات في اختزان المعلومات واسترجاعها، فقد رصدت الطبعة الثانية من كتاب لانكستر تطورات المجال حتى منتصف العقد الثامن من القرن نفسه، وواصلت الطبعة الثالثة من ذلك الكتاب الاهتمام بما شهده المجال من تطورات، حتى مطلع العقد الأخير من القرن العشرين، قبل أن تفجر ثورة الإنترنت شحنة التطورات المتلاحقة، التي لا نزال نخوض غمارها حتى اليوم فحسب، وإنما نستشرف أيضا آفاق مستقبلها وتداعياتها المرتقبة. ولقد كنت أترقب صدور الطبعة الرابعة من كتاب لانكستر، الذي طالما بهرنا بأدائه العلمي واللغوي المتميز. إلا أنه طال الانتظار، ولم تظهر الطبعة الرابعة من ذلك الكتاب. وأرى في الكتاب الذي نقدم لترجمته العربية اليوم، ضربا من التعويض الجزئي على الأقل.

فقد صدرت الطبعة الأولى من كتاب هتنج تشو **Heting Chu**، في مطلع القرن الحادي والعشرين (عام ٢٠٠١)، وترجمت إلى اللغة الكورية عام ٢٠٠٥، ونشرت في الهند عام ٢٠٠٩. ثم صدرت الطبعة الثانية من هذا الكتاب عام ٢٠١٠، متضمنة الكثير من الإضافات والتعديلات التي تتناول ما استجد من تطورات، وتستثمر رسائل التلقيم المرتد التي تضمنتها آراء النقاد، فيما حظيت به الطبعة الأولى من مراجعات. وقد ورد بيان الإضافات والتعديلات تفصيلا في تصدير هذه الطبعة الثانية.

ولدت مؤلفة هذا الكتاب عام ١٩٥٧، وحصلت على الليسانس في علم المكتبات، من جامعة بكين **Peking University** في الصين، وعلى الماجستير في علم المكتبات والمعلومات، من جامعة ماجل **McGill University** في كندا، وعلى الدكتوراه في دراسات المعلومات من جامعة دركزل **Drexel University** في الولايات المتحدة الأمريكية. ثم عملت تشو أستاذًا مساعدا، لمدة عام واحد في جامعة سانت جون **St. John's University**. وانضمت إلى هيئة التدريس فى جامعة لونج أيلاند **Long Island**، بالولايات المتحدة الأمريكية، في سبتمبر عام ١٩٩٤، حيث تعمل الآن أستاذا بمعهد بالمر لعلم المكتبات والمعلومات **Palmer School of Library and Information Science**. وقد ألفت تشو خمسة كتب بالصينية، حول استرجاع المعلومات، وما يتصل به من موضوعات. وتشمل اهتمامات تشو في البحث والتدريس، استخدام تقنيات المعلومات في المكتبات وتنظيم المعلومات واسترجاعها، ومناهج البحث، والاتصال العلمي، والقياسات الوراقية **Bibliometrics**، وقياسات النشاط العلمي **Scientometrics**. وتتراوح موضوعات بحوث تشو ما بين المكتبات الرقمية، والكتب والدوريات الإلكترونية، ومحركات البحث في العنكبوتية، وتطبيقات الجيل الثاني من العنكبوتية **Web O.2**.

ومنذ انضمامها إلى هيئة التدريس فى جامعة لونج أيلاند، تقوم تشو بتدريس مقرر واحد في المرحلة الجامعية الأولى، وثمانية مقررات في مستوى الماجستير، كما أدارت خمس قاعات للبحث للدكتوراه. وتشمل خدمات تشو في لونج أيلاند المستويات

المحلية والوطنية والدولية. وتشارك تشو في أعمال ثماني لجان على مستوى الجامعة، كما تشارك في تحكيم ما يقدم للنشر في ست دوريات متخصصة في علم المعلومات، فضلا عما يصدر عن دارين للنشر من كتب، كما أنها عضو في مجالس تحرير عدد من الدوريات المتخصصة في علم المعلومات. وفضلا عن كثافة الاستشهاد المرجعي بأعمالها، كثيرا ما تتلقى تشو طلبات الحصول على مستلات من مقالاتها، من أناس في مختلف أنحاء العالم.

هذا عن المؤلفة، فماذا عن الكتاب؟

يتكون هذا الكتاب من اثني عشر فصلا، يأتي أولها بمثابة تمهيد، يلقي نظرة شاملة على التطور التاريخي لتنظيم المعلومات واسترجاعها، ويعرف بعدد من رواد المجال، ويناقش المفاهيم الأساس، ويحلل مكونات المجال، ويعرض لأبرز مشكلاته. ويعالج الفصلان الثاني والثالث تنظيم المعلومات، موزعا على قسمين؛ يتناول أولهما (الفصل الثاني) الطرق الرئيسة للتنظيم، من التكشيف والتصنيف والتلخيص، بينما يتناول الثاني (الفصل الثالث) ما يتصل بالتنظيم من موضوعات، تبدأ بما وراء البيانات، وخصائص النصوص الكاملة ومشكلاتها، وتنظيم معلومات الوسائط المتعددة. وتستأثر اللغة، لما لها من دور رئيس في تنظيم المعلومات واسترجاعها، بالفصل الرابع، الذي يتناول اللغة الطبيعية، واللغات المقيدة، على نحو مقارن، يبرز ما لكل من الشكلين وما عليه، مع الاهتمام بخصائص لغة تنظيم المعلومات واسترجاعها في العصر الرقمي. وتتناول الفصول من الخامس حتى التاسع أساليب الاسترجاع،

وصياغة الاستفسارات. ويتناول الفصل السادس طرق الاسترجاع، كالبحث أو التنقيب والتصفح، والتكامل بين البحث والتصفح في الاسترجاع. ويناقش الفصل السابع نماذج استرجاع المعلومات، بدءا بالمضاهاة كأساس، ثم الجبر البوليني، وأنموذج الفضاء المتجهي، وأنموذج الاحتمالات، وامتدادات النماذج الرئيسة لاسترجاع المعلومات، واتجاهاتها الحديثة. ويتناول الفصل الثامن نظم استرجاع المعلومات في تسلسلها التاريخي، بدءا بنظم الخط المباشر، ونظم الأسطوانات المكتنزة، ونظم الفهارس المتاحة على الخط المباشر، ونظم استرجاع الإنترنت، ويختتم بالاتجاهات الحديثة في نظم استرجاع المعلومات. ويتناول الفصل التاسع استرجاع المعلومات المتفردة في محتواها أو في شكلها، كالمعلومات متعددة اللغات، والمعلومات متعددة الوسائط، ومعلومات النصوص الفائقة، ومعلومات الوسائط الفائقة.

ويحظى المستفيد من المعلومات بالاهتمام في الفصل العاشر، الذي يتناول المستفيدين واحتياجاتهم المعلوماتية، والأنموذج المعرفي، وغيره من النماذج التي تركز على المستفيد، والتفاعل بين المستفيد والنظام، ويختتم بإلقاء نظرة على المستفيد واسترجاع المعلومات في العصر الرقمي. وكما في أي نشاط، فإن للتقييم أهمية لا تنكر في تنظيم المعلومات واسترجاعها. ويتناول الفصل الحادي عشر التقييم، من حيث مقاييسه، ومعايير تقييم نظم استرجاع المعلومات بكل فئاتها، فضلا عن القابلية للاستخدام بوصفها معياراً للتقييم. ويفسح هذا الفصل المجال لمعالجة أكبر مشروعين للتقييم في تنظيم المعلومات واسترجاعها؛ وهما تجارب كرانفيلد، وسلسلة مؤتمرات استرجاع النصوص.

ويعالج الفصل الثاني عشر الأخير الذكاء الاصطناعي في تنظيم المعلومات واسترجاعها، ويبدأ بإلقاء نظرة على بحوث الذكاء الاصطناعي، ومعالجة اللغة الطبيعية، لما لها من دور محوري في تنظيم المعلومات واسترجاعها، ثم الجوانب التطبيقية لمعالجة اللغة الطبيعية، ثم العنكبوتية الدلالية من حيث هندستها المعمارية، وتطبيقاتها في استرجاع المعلومات، والتحديات التي تواجهها بوصفها أحد تطبيقات الذكاء الاصطناعي. ويختتم هذا الفصل بدور الذكاء الاصطناعي في تنظيم المعلومات واسترجاعها.

وفضلا عن الثراء المعلوماتي، تتسم المعالجة في هذا الكتاب بالتسلسل المنطقي القائم على التدرج من العام إلى الخاص، ومراعاة البعد التاريخي. وعلى هذا النحو يكتسب هذا الكتاب الطابع الدراسي، الذي يؤهله للارتباط بمقررات تنظيم المعلومات واسترجاعها في البيئة الرقمية.

وقد حرصت المؤلفة على ربط محتوى الكتاب برصيد الإنتاج الفكري في مجال تخصصه، بقوائم المراجع الملحقة بكل فصل. وقد بلغ عدد الأعمال المستشهد بها في الفصل الأول ٥٤ عملا، والأعمال المستشهد بها في الفصل الثاني ٢١ عملا، وفي الفصل الثالث ٢٠ عملا، وفي الفصل الرابع ٢٤ عملا، وفي الفصل الخامس سبعة عشر عملا، وفي الفصل السادس ثمانية عشر عملا، وفي الفصل السابع ٢٢ عملا، وفي الفصل الثامن ٤٩ عملا، وفي الفصل التاسع ٥٨ عملا، وفي الفصل العاشر ٥٣ عملا، وفي الفصل الحادي عشر ٩٩ عملا، وفي الفصل الثاني عشر ٤٥ عملا.

والإنجليزية هي اللغة الغالبة على الأعمال المستشهد بها. وتتراوح هذه الأعمال بين مقالات الدوريات، وتقارير البحوث، وبحوث المؤتمرات، ووثائق المواصفات المعيارية، والكتب وفصول الكتب. وتلتزم تشو بأسلوب الاستشهاد المرجعي المتبع في المراجعات العلمية؛ إذ تورد بيانات الأعمال المستشهد بها في نهاية كل فصل، مرتبة هجائيا بأسماء المؤلفين، متبوعة بتواريخ النشر. ويتم الربط بين قوائم المراجع والنص بتسجيل هذين العنصرين بين قوسين في المتن.

ولقد ترددت طويلا أمام مصطلح **Representation** في عنوان أصل هذا الكتاب؛ فلهذا المصطلح ثلاثة مقابلات محتملة في العربية؛ أولها، وربما كان أبسطها "تمثيل" وثانيها "التعبير عن" وثالثها "التنظيم". ويمكن لمصطلح "التمثيل" أن يبدو غامضا في هذا السياق. أما "التعبير عن" فيمكن أن يكون مرادفا للتمثيل، بينما تشعبت موضوعات هذا الكتاب على نحو يتجاوز المجال الدلالي لكل من "التمثيل" و"التعبير عن"، ويربطنا بالمجال الدلالي لمصطلح "التنظيم"، ذلك المصطلح الذي ارتبط بالاسترجاع ارتباطا وثيقا. وقد فضلنا هذا المصطلح الأخير، الشامل الذي يغطي الجوانب اللغوية والتقنية والتنظيمية والمسلكية للتعامل مع المعلومات. يضاف إلى ذلك أن استعمال "استرجاع المعلومات" عنواناً للكتاب في هذا المجال، يعد كافيا في حد ذاته، لأنه دائما ما يكون هناك وراء الاسترجاع الفعال تنظيم واختزان.

ومن الطبيعي لكتاب يعالج أحدث تطورات تنظم المعلومات واسترجاعها، أن يتضمن عددا كبيرا من المصطلحات التي تدل على ما استجد من مفاهيم، أو تمثل

وسيمات جديدة لمفاهيم قديمة. وقد بذلت قصارى جهدي في اختيار مقابلات عربية لهذه المصطلحات. وما هذه المقابلات العربية إلا اجتهادات، يمكن أن تصيب أو تخطئ. ونحمد الله تعالى أن هدانا للصواب، ونعتذر عن أي خطأ لم نقصده. وقد حرصت على تسجيل المصطلحات التي وردت في ثنايا الكتاب، ومقابلاتها العربية المقترحة، في ثبت ورد في نهاية هذا الكتاب.

وإذ أقدم ناتج جهدي في ترجمة هذا الكتاب، أرجو أن أكون قد وفقت في نقل رسالة المؤلفة كاملة غير منقوصة؛ فقد حرصت قدر طاقتي على مراعاة الأمانة، بالتوفيق قدر الإمكان بين نظامين لغويين، لكل منهما بيئته الاجتماعية والثقافية، وإمكاناته وخصائصه النحوية والدلالية. وبصدور هذا الكتاب تكتمل بحول الله تعالى، رباعية ترصد تطورات مجال استرجاع المعلومات على مدى حوالي نصف قرن. وأسأل الله تعالى أن يكون جهدي هذا خالصا لوجهه سبحانه، وأن يكون نافعا فيما قصد به، إنه ولي ذلك والقادر عليه، وآخر دعوانا أن الحمد لله رب العالمين.

أ.د. حشمت قاسم

القاهرة

جمادى الآخرة ١٤٣٥هـ/ أبريل، ٢٠١٤م

تصدير الطبعة الثانية

ما يقرب من ثماني سنوات منذ انتهيت من تأليف الطبعة الأولى من هذا الكتاب، في يوليو عام ٢٠٠١. وتدل الفترة الزمنية في حد ذاتها على الحاجة إلى طبعة جديدة، في مجال تنظيم المعلومات واسترجاعها، سريع التطور. يضاف إلى ذلك صدور أكثر من عشرين مراجعة على الأقل، للطبعة السابقة من هذا الكتاب. وتشتمل تلك المراجعات على كثير من الاقتراحات والتعليقات القيمة، التي لا أقدرها عاليا فحسب، وإنما يمكن أن أود أيضا تضمينها في هذا الكتاب. وقد أصبح هذان السببان مجتمعين، فيما بعد العامل المحفز لتأليف هذه الطبعة الثانية.

وبينما تم تحديث الكتاب كاملا ومراجعته ، تشتمل القائمة التالية على بيانات التعديلات الجوهرية التي أدخلتها على كل فصل في الطبعة الحالية:

- الفصل الأول: إضافة قسم جديد؛ ١/ ١/ ٢/ ٥ عن كارن سبارك جونز **Karen Spärck Jones (1935-2007)**.
- الفصل الثاني: إضافة قسمين جديدين؛ ٢/ ١/ ٤ التمييز الاجتماعي **Social Tagging**، و٢/ ٣/ ١/ ٤ محتوى المواقع الثرية **RSS**.
- الفصل الثالث: إضافة قسم جديد؛ ٣/ ١/ ٣/ ٣ محددات هوية الكيانات الرقمية **Digital Object Identifiers**.

• الفصل الرابع: إضافة ثلاثة أقسام؛ ٤/ ٤/ ١ التقسيمات الهرمية Taxonomies، و٤/ ٤/ ٢ التصنيفات المتعارف عليها Folksonomies، و٤/ ٤/ ٣ والتصنيفات التخصصية Ontologies.

• الفصل الخامس: إضافة عملية البحث / استرجاع خدمات العنكبوتية Search/Retrive Web Service (SRW)، والبحث / الاسترجاع عن طريق المحدد الموحد لمكان المصدر Search / Retrieve via URL (SRU)، إلى القسم ٥/ ١/ ٢/ ٤ البحث في مراصد بيانات متعددة Multiple Databare Searching.

• الفصل الثامن: إضافة الجيل التالي من الفهارس المتاحة على الخط المباشر OPACs، إلى القسم ٨/ ٣ الفهارس المتاحة على الخط المباشر OPACs: الفهارس الإلكترونية للمكتبات بوصفها أنظمة لاسترجاع المعلومات، وإدخال تعديلات جوهرية على القسم ٨/ ٤/ ٢/ ٤ أساليب الترتيب الطبقي، وإضافة قسم جديد ٨/ ٥/ ٢ العنكبوتية ٢.٠ ونظم استرجاع المعلومات.

• الفصل التاسع: تعديلات جوهرية في القسم ٩/ ١ المعلومات متعددة اللغات، والقسم ٩/ ٢/ ٢ استرجاع الأصوات، والقسم ٩/ ٢/ ٣ استرجاع الصور المتحركة.

• الفصل العاشر: إضافة قسم جديد ١٠/ ٢/ ٢ النماذج الأخرى لاسترجاع المعلومات التي تركز على المستفيد، وإضافة واجهة المستفيد العضوية organic user interface (OUI) إلى القسم ١٠/ ٣/ ١/ ٤ نماذج أخرى لتفاعل المستفيد مع النظام.

• الفصل الحادي عشر: حذف القسم ١١/ ٢/ ٢ معايير تقييم نظم الأسطوانات الضوئية المكتنزة **Evaluation criteria for CD-ROM systems**، وإضافة القسمين ١١/ ٢/ ٤ معايير تقييم نظم الاسترجاع متعددة الوسائط، و١١/ ٢/ ٥ القابلية للاستخدام بوصفها معيارًا للتقييم، وإدخال تعديلات جوهرية في القسم ١١/ ٣/ ٢/ ١/ ٤ مهام الاسترجاع.

• الفصل الثاني عشر: التوسع وإجراء مراجعات جوهرية، تشمل حذف القسم ١٢/ ٢/ ٢ أنموذج اللغة الطبيعية، وإضافة أقسام جديدة؛ ١٢/ ٢/ ٢ التلخيص الآلي، و١٢/ ٢/ ٣ الرد على الاستفسارات، و١٢/ ٢/ ٤ البحث في اللغة الطبيعية، و١٢/ ٣ العنكبوتية الدلالية.

وفضلا عن ذلك، أود أن أوضح أن بؤرة اهتمام هذا الكتاب وتوجهه يستندان إلى ما تلقيت من تلقيم مرتد **feedback**، حول الطبعة السابقة. ونتناول تنظيم المعلومات واسترجاعها **IRR**، بادئ ذي بدء، من منظور علم المكتبات والمعلومات، لا من وجهة نظر علوم الحاسب. ومن ثم فإن دقائق تصميم النظم وتنفيذها، كالخوارزميات **algorithms** على سبيل المثال، لم يتضمنها هذا الكتاب. وثاني، تمت معالجة طرق تنظيم المعلومات بوصفها مكونا لا غنى عنه في عملية استرجاع المعلومات؛ لا بوصفها مجالا موازيا لاسترجاع المعلومات. ولهذا، فإن تغطية طرق تنظيم المعلومات في هذا الكتاب، أقل حجماً من تغطية أساليب استرجاع المعلومات. ثالثا، لم تحظ موضوعات معينة (كسلوكيات البحث عن المعلومات مثلا) إلا بمعالجة

موجزة في هذا الكتاب، وذلك لسببين: (١) لم يقصد بهذا الكتاب معالجة هذه الموضوعات باستفاضة، (٢) يمكن الحصول على تغطية ممتازة لهذه الموضوعات في مصادر أخرى، وردت على وجه التحديد بوصفها مراجع في نهاية الفصول المتصلة بها.

وأود أخيرا شكر جامعة لونج أيلاند **Long Island University** لمنحي إجازة تفرغ، ما كان من الممكن لي بدونها أن أتمكن من تأليف طبعة جديدة من هذا الكتاب. كما أني أعترف بكل التقدير للجهود التي بذلها مساعدي، طالب الدراسات العليا فنفاي أويانج **Fenfei Ouyang** في تجميع المادة لي. وفضلا عن ذلك فإنني أقدر فعلا مشورة وتوجيه سامانتا هيستنجز **Samantha Hastings**، محرر سلسلة كتب الجمعية الأمريكية لعلم وتقنيات المعلومات **ASIST Monograph Series Editor**، وآمي ريف **Amy Reeve**، المحرر التنفيذي للكتب، وجون بي. بريانز **John B. Bryans**، رئيس التحرير والناشر، بمؤسسة حاضر المعلومات **Information Today Inc.**، في أثناء إعداد الطبعة الحالية. وإنها لمتعة فعلا أن أعمل مع كل هؤلاء جميعا كما هو الحال دائما.

هتنج تشو

لونج أيلاند، نيويورك

تصدير الطبعة الأولى

كتاب آخر حول استرجاع المعلومات؟ نعم، لأن هناك موضوعات وتطورات جديدة بحاجة لأن تعالج في هذا المجال، ونحن على مشارف العصر الرقمي. يضاف إلى ذلك تكريس فصلين كاملين لطرق تنظيم المعلومات، كخطوة ينبغي اتخاذها لكي تصبح المعلومات قابلة للاسترجاع.

ونبدأ بنظرة شاملة على طرق تنظيم المعلومات واسترجاعها، في الفصل الأول، الذي يلقي نظرة على المفاهيم الأساس، والأعلام الأساس، والأحداث الأساس، ومراحل التطور الجوهرية للمجال. ويتناول الفصلان الثاني والثالث المقاربات الأساس لطرق تنظيم المعلومات، والموضوعات الأخرى المتصلة بذلك. ونظراً لأهمية اللغة في تنظيم المعلومات واسترجاعها؛ يناقش الفصل الرابع اللغة الطبيعية واللغات المقيدة، نوعي اللغة المستخدمين في المجال، بالإضافة إلى علاقات هذين النوعين وخصائصهما. وتركز الفصول الخامس والسادس والسابع على مختلف أوجه الاسترجاع؛ إجراءات الاسترجاع، ومقاربات الاسترجاع، ونماذج الاسترجاع. ثم نتناول بعد ذلك الفئات الرئيسة لنظم استرجاع المعلومات، وذلك في الفصل الثامن، مع اهتمام خاص بنظم استرجاع الإنترنت، النجم البازغ في أسرة نظم استرجاع المعلومات. ويستكشف الفصل التاسع استرجاع المعلومات متعددة اللغات، ومعلومات الوسائط المتعددة، والمعلومات فائقة التنظيم **hyperstructured**.

ويحظى البعد الخاص بالمستفيد، وهو جانب جوهري في تنظيم المعلومات واسترجاعها، بالتغطية في الفصل العاشر. ويستعرض الفصل الحادي عشر قضية التقييم المعقدة، متعددة الجوانب في المجال، ويشمل مقاييس التقييم، وطرق التقييم، ومشروعات التقييم الكبرى. ويحلل آخر فصول هذا الكتاب، وهو الفصل الثاني عشر، دور الذكاء الاصطناعي واحتمالاته في تنظيم المعلومات واسترجاعها.

ولقد حاولت تقديم نظرة منطقية شاملة، وإن كانت غير تقنية للمجال، باستعمال لغة جلية بسيطة، حيثما أمكن ذلك، لشرح موضوعات معقدة. وينصب تركيز هذا الكتاب على مبادئ وأسس تنظيم المعلومات واسترجاعها، لا على وصف إجراءات أو نظم أو ممارسات بعينها في المجال. وبمجرد أن يلم القارئ بمبادئ وأسس تنظيم المعلومات واسترجاعها هذه، فإنه يمكن أن يصبح قادرا على تطبيق هذه المبادئ والأسس في مختلف المواقف والبيئات. كذلك نولي الاهتمام، على وجه التحديد، للموضوعات والتطورات فيما يتعلق بتنظيم المعلومات واسترجاعها في العصر الرقمي.

وبينما يتوجه "تنظيم المعلومات واسترجاعها" نحو المستفيد، فإنه يمكن لمصممي نظم معالجة المعلومات واسترجاعها، أن يجدوا هذا الكتاب مفيدا بالنسبة للإلمام بالمجال من منظور المستفيدين. وأهدف هنا إلى كتاب يمكن أن يكون مفيدا لكل من يهتم بالإحاطة بالمجال، وخصوصا أولئك الوافدين إليه حديثا. وأوصي بقوة بأن يقرأ المبتدئون الكتاب بالترتيب الذي ورد به محتواه، نظراً لأن الفصول المتأخرة تستند إلى الفصول المتقدمة.

وبنظرة إلى ما مضى، فإنه ربما لم يكن لي أن أؤلف هذا الكتاب، في هذه المرحلة من حياتي المهنية، ما لم يبادر زميلي ريتشارد سميراجليا **Richard Smiraglia**،

بالاتصال المباشر من أجلي. وأعرب عن امتناني أيضا لتلاميذي الذين شاركوني اهتمامي بالمجال وحماسي له. كما أقدر أيضا إجازة التفرغ التي منحتني إياها جامعة لونج أيلاند؛ فما كان من الممكن بدون هذه الإجازة أن تتاح فسحة الوقت اللازم لإنجاز هذا المشروع. كما أود أن أشكر مايكل كونيج **Michael Koenig**، عميد معهد بالمر **Palmer School** الذي أنتسب إليه، وجون بريانز **John Bryans**، رئيس تحرير قسم نشر الكتب بمؤسسة حاضر المعلومات **Information Today Inc.**؛ لأنهما جعلا نشر مسودتي حقيقة واقعة.

ولقد كان هذا الكتاب، وإلى حد بعيد، جزءا لا يتجزأ من حياتي العائلية، في أثناء التأليف؛ لأن زوجي وابنتي كانا أيضا منخرطين بعمق في عملي، إذ أصبح هذا الكتاب، على سبيل المثال، موضوعا منتظما للحديث على مائدة العشاء، وغالبا ما كانت أسرتي تقضي نهاية الأسبوع، وساعات المساء بدوني في المنزل. وربما كانت ابنتي فانجفاي **Fangfei** أكثر تعرضا "لاسترجاع المعلومات" من أي أطفال آخرين في سنها؛ فقد كانت تحثني باستمرار على العمل في كتابي، على الرغم من أنها كانت في قرارة نفسها، ربما تفضل أن أقضي الوقت معها. ولقد كان ما حظيت به من حب وتفاهم ودعم من أسرتي، مصدرا للطاقة والإلهام على الدوام.

هتنج تشو

لونج أيلاند، نيويورك

الفصل الأول

تنظيم المعلومات واسترجاعها؛ نظرة عامة

يرجع تنظيم المعلومات واسترجاعها، الذي يعرف أيضا بالاستخلاص والتكشيف والبحث عن المعلومات ومعالجة المعلومات وإدارتها، إلى النصف الثاني من القرن التاسع عشر، حينما وضعت خطط لتنظيم المعرفة والوصول إليها (كتصنيف ديوي العشري على سبيل المثال عام ١٨٧٦) *(Wynar, 1980)*. إلا أن البحث في تنظيم المعلومات واسترجاعها لم يصبح مجالا رئيسا في علم المعلومات، إلا بعد نشوب الحرب العالمية الثانية. ومنذ ذلك الحين تتجه عقول نيرة من مختلف المجالات، نحو هذا المجال الناشئ، بينما تستخدم تقنيات المعلومات بدرجات متفاوتة من التقدم والنضج، لتيسير البحث والتطوير في تنظيم المعلومات واسترجاعها.

ويتسم تاريخ تنظيم المعلومات واسترجاعها هنا وتطوره، بالتركيز على الملامح الرئيسة لكل فترة زمنية، والرواد البارزين في المجال. ومن ثم فإن المفاهيم الرئيسة التي يتضمنها هذا الكتاب تحظى بالمعالجة، مشفوعة بمناقشة للمكونات الجوهرية لتنظيم المعلومات واسترجاعها. ونختتم هذا الفصل بمحاولة استكشاف المشكلة الجوهرية للمجال، وهي كيف يتم الحصول على المعلومات المناسبة للمستفيد المناسب في الوقت المناسب.

تاريخ وتطور تنظيم المعلومات واسترجاعها:

تاريخ تنظيم المعلومات واسترجاعها ليس بالبعيد. وتحدد النظرة الراجعة للمجال معالم الطلب المتزايد، والنمو السريع، ومرحلة إزالة الغموض *demystification*، وحقبة المشابكة، بوصفها المراحل الأربع الرئيسة التي مر بها تنظيم المعلومات واسترجاعها في تطوره.

المراحل الرئيسة:

الطلب المتزايد (أربعينيات القرن العشرين - مطلع خمسينيات القرن نفسه):

شهدت الحرب العالمية الثانية الميلاد الرسمي لمجال تنظيم المعلومات واسترجاعها؛ فنظرا لظروف الحرب، تم إنتاج عدد هائل من التقارير التقنية والوثائق التي تسجل وتوثق أنشطة البحث العلمي والتطوير، المرتبطة بإنتاج العتاد الحربي. ولا يمكن للإنسان مواجهة مثل هذه المهمة الضخمة لتنظيم المعلومات واسترجاعها، دون النظر في الجوانب الأخرى لمعالجة المعلومات وإدارتها، كالانتقاء، والبث والحفظ. وكما ذهب بوش *Bush (1945)*:

> هناك جبل نام من البحوث، إلا أن هناك دليلا متزايدا على أننا نغوص اليوم في مستنقع، نظرا لتزايد التخصص. ويترنح الباحث تحت وطأة نتائج وخلاصات جهود الآلاف من الباحثين الآخرين، تلك الخلاصات التي لا يتسع وقت الباحث لاستيعابها، وناهيك عن تذكرها، عندما تظهر *(P.101)*.

حقـا، لقـد أصـبحت الحاجـة إلى تـسجيل الكميـات الهائلـة مـن المعلومـات وتنظيمها واضحة تماما، وضاغطة أيضاً؛ ففي مجالات الفيزياء والأحيـاء عـلى سـبيل المثال:

> يقدر ما ينشر في العلوم البيوطبية *Biomedical* على سـبيل المثـال، بمليـوني بحـث سـنويا *(McCandless, Skweir & Gordon, 1964, P. 147)*. ويمكن الاطلاع على هـذه البحـوث بمعدل اثنين في الساعة، إذا سـلمنا بـأن القـارئ يقـظ، وبإمكانـه القراءة بحوالي سبعين لغة، والوثـائق في متنـاول يديـه. وإذا كـان الاطلاع على الدوريات يقتصر على سـاعة واحـدة يوميـا، و٣٦٥ يوما في العام، فـإن الاطـلاع عـلى إنتـاج عـام واحـد مـن النتـاج الفكري العالمي في العلوم البيوطبية، يمكن أن يستغرق أكثـر مـن ٢٧.٤ قرنا *(Borko & Bernier, 1975, p. 6)*.

وعلى الـرغم مـن أنـه قـد لا يكـون مـن الممكـن التعـرف عـلى الكـم الفعـلي للمعلومات التقنية، التـي تـم إنتاجهـا في العقـدين الخـامس والـسادس مـن القـرن العشرين، على وجه الدقة، فإنه من الممكن تقدير كـم هـذه المعلومـات، وفقـا لمجـال العلوم البيوطبية الذي سـبقت الإشـارة إليـه. ولم يعـد بإمكـان البـشر الاعـتماد عـلى مهاراتهم وذاكرتهم أو إمكانات الملفـات الشخـصية دون سـواها، مـن أجـل تنظـيم المعلومات واسترجاعها بكفاءة، كلـما دعـت الحاجـة إلى ذلـك. وإنـما تطلـب الأمـر الاعتماد على الجهود الجماعية، في مجال تنظيم المعلومات واسـترجاعها، ممـا أسـفر عـن

النظم المصممة خصوصًا لذلك الغرض، حتى وإن كانت نظمًا يدوية، كالكشافات الترابطية *coordinated indexes* التي ظهرت عام ١٩٥١ *(Gull, 1956)*.

النمو السريع (من خمسينيات القرن العشرين حتى ثمانينياته):

تمثل عقود هذه الفترة السنوات الذهبية في تطور تنظيم المعلومات واسترجاعها؛ فقد دخلت الحاسبات رسميا إلى المجال فيما بين عامي ١٩٥٧ و١٩٥٩، عندما استخدم هانز بيتر لون *Hans Peter Luhn* حاسبا لا لمعالجة مضاهاة الكلمات المفتاحية ومهام الفرز فحسب، وإنما للعمل الفكري المتصل بتحليل مضمون النصوص التحريرية أيضا *(Salton, 1987)*.

ويدل ظهور نظم الخط المباشر *online* مثل ديالوج *DIALOG* في ستينيات وسبعينيات القرن العشرين، على التحول من الاسترجاع اليدوي إلى الاسترجاع الإلكتروني للمعلومات. وقد وصف هان *Hahn (1996)* نظم الخط المباشر الرائدة، التي تطورت في ذلك الوقت قائلا:

> لقد كانت تتمتع ببعض الملامح المتطورة البارزة، كمكانز الخط المباشر والمخرجات المرتبة طبقيا، وإدخال المترادفات، تلقائيا في صيغ أو استراتيجيات البحث، والمنطق البوليني، وبتر كواسع الكلمات وصدورها، والبحث في المراجع المستشهد بها، وبحث النصوص الحرة للغة الطبيعية. وكانت بعض النظم تتضمن برامج لتجميع البيانات لمراقبة الإفادة من النظم ومدى الرضا عنها *(P. 34)*.

وإلى جانب نمو نظم الخط المباشر ونضجها، تم تطوير أساليب آلية وتلقائية لتنظيم المعلومات واسترجاعها، وإجراء التجارب على هذه الأساليب، مدعومة بالتطورات التي شهدتها تقنيات استخدام الحاسبات أو الحوسبة. وكرست عناصر بشرية من مختلف المجالات، وخصوصا علوم الحاسب، جهودها للبحث والتطوير في هذا المجال، إلا أنه كانت هناك مشكلات أخرى تنتظر البحث، كما أوجز سالتون *Salton (1987)* في أحد أعماله:

> على الرغم من أن قدراً كبيرا من التقدم قد تحقق، خلال السنوات الثلاثين الماضية، في معالجة النصوص واسترجاع المعلومات، وخصوصا في قطاعات تحرير النصوص وإنتاج الوثائق، وتحديد أو تعيين المصطلحات الكشفية، والتجميع الديناميكي لعمليات البحث وصياغة الاستفسارات، فإنه لا يبدو هناك سوى تطورات جوهرية محدودة في مجالات الاستيعاب الحقيقي للنصوص *(P. 379)*.

مرحلة إزالة الغموض (ثمانينيات القرن العشرين وتسعينياته):

على الرغم من أن نظم الخط المباشر التي سبقت الإشارة إليها، كانت مصممة للمستفيدين ذوي الفئات المختلفة من الاحتياجات المعلوماتية، فإن تلك النظم لم تكن مصممة على النحو الذي يكفل للمستفيدين إجراء عمليات البحث فيها مباشرة، دون أي تدريب أو مساعدة من اختصاصيي المعلومات. وبعبارة أخرى فإنه لم يكن هناك

سوى الوسطاء كاختصاصيي المكتبات وغيرهم من اختصاصيي المعلومات، من يستطيعون إنجاز مهام البحث نيابة عن المستفيدين. وفضلا عن ذلك كانت الإفادة من مثل هذه النظم باهظة التكلفة، من أجل الوصول إلى المعلومات، نظرا لتقاضي مجموعة ضخمة من الرسوم (كرسوم الاتصالات بعيدة المدى، والربط، ورسوم مراصد البيانات، على سبيل المثال) وذلك مقابل عملية بحث واحدة. وكان مصطلح المستفيد النهائي يدل حينئذ على أولئك الذين يحتاجون إلى نوع ما من المعلومات، ولكنهم لم يكن بإمكانهم إجراء عمليات البحث على الخط المباشر فعلا بأنفسهم. وقد تغير معنى مصطلح المستفيد النهائي تدريجيا، عندما بدأ استخدام الحاسبات الشخصية في استرجاع المعلومات، وعندما بدأت نظم الأسطوانات المكتنزة *CD-ROM*، والفهارس المتاحة على الخط المباشر *(OPACs)*، تطبق في منتصف ثمانينيات القرن العشرين.

لقد كان التعامل مع نظم استرجاع المعلومات على الخط المباشر يتم ، فيما مضى عن طريق وسائل متنوعة، كالمنافذ الطابعة *printer terminal*، ومنافذ أنبوبة أشعة المهبط *(CRT)*. ومن نافلة القول أن التفاعل بين من يقوم بإجراء عمليات البحث من جهة، والنظام من جهة أخرى، لم يكن جاذبا أو متعاطفا أو ميسرًا. وعندما دخلت الحاسبات الشخصية إلى استرجاع المعلومات، تبين للمستفيد النهائي أن عملية الاسترجاع كانت تبدو أقل رهبة مما كانت عليه بكثير؛ نظرا لأنه أصبح من الممكن إجراء شكل ما من "المحادثة" الودية، أو التفاعل بين المستفيد والنظام.

وقد أدى تطبيق نظم الأسطوانات المكتنزة *CD-ROM*، والفهارس المتاحة على الخط المباشر *OPACs*، إلى أن يصبح من الممكن للمستفيد النهائي إجراء عمليات البحث لنفسه، وذلك بالقضاء على غموض نظم استرجاع المعلومات الذي كان قائما من قبل، في شكل الخط المباشر فقط، كما كان هؤلاء المستفيدون موزعين جغرافيا في أماكن نائية، ولم يعودوا بحاجة لأن يحفلوا بتكلفة الخط المباشر، عندما يقومون بالبحث، وذلك في نظم الأسطوانات المكتنزة، والفهارس المتاحة للجمهور على الخط المباشر. ومنذ ذلك الحين أصبحت نظم استرجاع المعلومات، وعلى نحو متزايد، نظما مصممة للمستفيدين النهائيين ولتستخدم من جانبهم.

حقبة المشابكة (منذ تسعينيات القرن العشرين حتى الآن):

لقد كان استرجاع المعلومات حتى ذلك الوقت نشاطا مركزيا، مما يعني أن مراصد بيانات نظم استرجاع المعلومات كانت تدار ماديا في موقع مركزي واحد. وإذا كان هناك من يرغبون في الإفادة من عدة نظم، كان عليهم إقامة علاقات مع نظم الاسترجاع المعنية كل على حدة. وفي مقابل ذلك، يكفل البحث اللامركزي للمستفيدين، الوصول إلى معلومات مراصد البيانات، باستخدام البنية الأساس للمشابكة *network infrastructure*. ولم تعد نظم استرجاع المعلومات تقتصر على منطقة جغرافية واحدة. وقد جعل ظهور الإنترنت من استرجاع المعلومات التشابكي حقيقة واقعة، بتوفير البنية الأساس اللازمة لهذا التطبيق.

وبالإضافة إلى مقومات البحث اللامركزي، أدت الإنترنت إلى إعادة تحديد معالم مجال تنظيم المعلومات واسترجاعها، فلم يحدث من قبل، في تاريخ تنظيم المعلومات واسترجاعها، أن استخدمت الكلمات المفتاحية الإحصائية، وغيرها من

الطرق المماثلة، بهذا القدر من الكثافة، في مثل هذا القدر الهائل من التنظيم الفائق *hyperstructure* ومعلومات الوسائط المتعددة. ولم يحدث مطلقا من قبل في تاريخ استرجاع المعلومات أن قام كثير من المستفيدين بإجراء عمليات البحث، على الخط المباشر دون مساعدة الوسطاء. ونتيجة لذلك يبدو طابع تنظيم المعلومات في هذه البيئة، مختلطاً إلى حد أن مصطلح الأزمة المنظمة *organized chaos* قد تم سكه خصوصًا لوصف الوضع الراهن. ومن ناحية أخرى، أصبح استرجاع النصوص الكاملة هو القاعدة وليس الاستثناء، على الإنترنت. وتتاح الآن أساليب الاسترجاع التي كانت فيما مضى تقتصر على مختبرات التجارب، وتتاح الآن بسهولة في نظم استرجاع الإنترنت، كألتافيستا *Alta Vista* وجوجل *Google*. وتطبق الآن النتائج العامة للبحوث التي يمكن الحصول عليها من البيئات التجريبية المنضبطة، على نطاق واسع، لتنظيم المعلومات واسترجاعها على الإنترنت.

وتتمثل حقبة المشابكة في الإنترنت، التي توفر منصة جديدة، ونافذة عرض لتنظيم المعلومات واسترجاعها في العصر الرقمي.

رواد المجال:

اجتذب مجال تنظيم المعلومات واسترجاعها كثيرا من الموهوبين المهتمين، في السنوات الخمسين الماضية، إلى حد أنه لا يمكن عمليا الإحاطة بكل هؤلاء في هذا المقام. إلا أن ما نورد فيما بعد من أعلام، يستحقون المعالجة كل على حدة في هذا الكتاب؛ لأن إسهاماتهم في المجال جد عظيمة. والمعيار الآخر لاختيار رواد استرجاع المعلومات في هذا القسم، أن جميعهم قد اختتموا حياتهم الأكاديمية وقت تأليف هذا الكتاب.

مورتيمر تاوبه (١٩١٠-١٩٦٥):

حصل مورتيمر تاوبه *Mortimer Taube* على شهادته للدكتوراه في الفلسفة، بجامعة كاليفورنيا في بركلي *University of California at Berkely*. وقد عمل اختصاصيا للمكتبات، في الإعارة، والفهرسة، والتزويد قبل أن يشغل منصبا في مكتبة الكونجرس، ليصبح مساعدا رئيسا للمراجع العامة والوراقة *bibliography* عام ١٩٤٥ *(Shera, 1978)*. وفي عام ١٩٥٢ أنشأ تاوبه مؤسسة الوثاقة *Documentation Inc.*، حيث بدأ هو وزملاؤه استكشاف طرق جديدة لتكشيف المعلومات واسترجاعها، بناء على عقد مع جهاز المعلومات التقنية للقوات المسلحة الأمريكية *U.S. Armed Services (Smith, 1993) Technical Information Agency*.

وقد أصبح النهج الجديد للتكشيف والبحث، يعرف في النهاية بالتكشيف الترابطي *coordinate indexing*. وقدم تاوبه مع ألبرتو طومسون *Alberto F. Thompson*، تقريراً بعنوان "التكشيف الترابطي للمجالات العلمية *The Coordinate Indexing of Scientific Fields*، أمام قسم الإنتاج الفكري الكيميائي بالجمعية الكيميائية الأمريكية *American Chemical Society*، في سياق مؤتمرها حول الوسائل الميكانيكية للتوثيق الكيميائي *Symposium on Mechanical Aids to Chemical Documentation*. ولم ينشر هذا التقرير مطلقا في دورية أو كتاب، إلا أن جال *Gull (1987)* ضمنه فيما بعد في أحد أعماله كملحق.

وكانت الحاجة إلى طرق جديدة للتكشيف والاسترجاع في ذلك الوقت مزدوجة؛ فقد أثبت أولا العدد الهائل من التقارير التقنية وغيرها من فئات الإنتاج الفكري العلمي، الذي أسفرت عنه البحوث التي أجريت في غضون الحرب العالمية الثانية، عدم ملاءمة نظم التكشيف والاسترجاع القائمة، التي كانت يدوية في المقام الأول. ثانيا، لم تكن الطريقتان الراسختان لتنظيم المعلومات، وهما الطريقة الهجائية، والطريقة الهرمية (أي رؤوس الموضوعات وخطط التصنيف) قادرتين على استيعاب المجالات التخصصية الجديدة، والتقنيات الجديدة، والمصطلحات الجديدة، التي أسفرت عنها جهود البحث والتطوير المتصلة بالحرب العالمية الثانية *(Smith, 1993)*. وحدث في هذه الظروف على وجه التحديد، أن اقترح تاوبه وطومسون، عام ١٩٥٢، التكشيف الترابطي *(Gull, 1987)*.

وباختلافه عن الطرق التقليدية للتكشيف والبحث، يقوم التكشيف الترابطي على استخدام المصطلحات الأحادية *Uniterms*، وتطبيق المنطق البوليني في استرجاع المعلومات. والمصطلحات الأحادية مصطلحات مفردة، يختارها المكشفون للدلالة على الجوانب المختلفة للوثائق. ويمكن النظر إلى المصطلحات الأحادية، على نحو ما، بوصفها الكلمات المفتاحية لأيامنا هذه؛ نظرا لأن كليهما يشتق من الوثائق الأصلية، ولا يبذل أي جهد من أجل التحكم في المصطلحات (كمراجعة المترادفات والمشترك اللفظي *homographs* على سبيل المثال). وعادة ما يتم استعمال عدة مصطلحات أحادية للدلالة على محتوى وثيقة ما، كما هو الحال بالنسبة لتكشيف الكلمات المفتاحية.

أما المنطق البوليني *Boolean logic*، وهو أحد فروع الفلسفة، فقد طوره جورج بول *George Boole* عام ١٨٤٩، بناء على تحليلاته لعمليات البرهنة *reasoning* البشرية، والقوانين الأساس التي تحكم العمليات العقلية *(Smith, 1993)*. وكانت عمليات البرهنة في نظر بول، إما إضافة مفاهيم أو فئات مختلفة من الأشياء، لتكوين مفاهيم أكثر تعقدا وإما تحليل مفاهيم مركبة إلى مفاهيم تحليلية مفردة أكثر بساطة *(Boole, 1854)*. ويتم التعبير عن العملية الأولى إيجازا بمعامل و *AND operator*، بينما تنطوي العملية الأخرى على معاملي أو *OR* وفيما عدا *NOT*. وبعد قرن على وجه التقريب، أدخل تاوبه هذه المبادئ إلى مجال تنظيم المعلومات واسترجاعها، في شكل التكشيف الترابطي.

وقد أرست جهود تاوبه التراكمية في التكشيف الترابطي، الأساس بالنسبة للبحث البوليني في البيئة المعتمدة على الحاسبات. وإذا ما أمكن تقسيم التخصصات الموضوعية إلى مفاهيم مفردة، يتم التعبير عنها بالمصطلحات الأحادية، فإنه يمكن استخدام الحاسبات لتنظيم المعلومات التي يتم التعبير عنها بهذا الشكل، والبحث عنها. وقد أفضت هذه النظرة الثاقبة في النهاية، إلى تطوير مختلف نظم الاسترجاع التي تقوم بإجراء جميع فئات البحث البوليني، الموضوع الذي نستكشفه بمزيد من التفصيل في أجزاء أخرى من هذا الكتاب. وفي الوقت نفسه، أصبحت الكشافات التي يتم تطويرها باستخدام طريقة الربط، تعرف بالكشافات الترابطية. وقد سميت عملية البحث في هذه الكشافات، بالربط بين الأفكار للعثور على المعلومات التي تدعو الحاجة إليها بربط المفاهيم *concept coordination* *(Smith, 1993)*.

بيد أن عبارة التكشيف الترابطي تسمية غير دقيقة *(Gull, 1987)*. وللمزيد من الدقة فإنه ينبغي أن يسمى التكشيف والبحث الترابطي، أو بمصطلحات أيامنا هـذه، التنظيم والاسترجاع الترابطـي *coordinate representation and retrieval* نظـرا لأنه لا يبدو مجرد طريقة للتكشيف، وإنما يستعمل للبحث أيضا. يـضاف إلى ذلـك أن تركيز هذه الطريقة كان ينصب على التحليـل أو التحليـل والتركيـب *synthesis*، ممـا حدا ببعض النقاد (مثل جال *Gull, 1987* وبـاو *Pao, 1989*) للتـساؤل مـا إذا كـان التكشيف الترابطي يربط بين الكلمات أم المفاهيم أو الأفكار. كـذلك كانـت مـشكلة الربط المزيف سببا في المخاوف؛ نظرا لأنه لم تكن هناك تقنية في هذه الطريقة تحول دون حدوث السقط(*) *false drops*. فإذا كان موضوع البحث المـراد، عـلى سـبيل المثـال، أسطوانة الحاسب *computer desk*، فإنه يمكن للنتـائج المـسترجعة أن تـشتمل عـلى وثائق تتـصل بأسـطوانة الحاسـب بالإضـافة إلى حاسـب المكتـب *desk computer*، وعبارات أخرى ترد بها كلمتا حاسب *computer* ومكتـب *desk* (كحاسـب سـطح المكتب *desktop computer* على سبيل المثال). ولما كانت المصطلحات الأحادية يـتم تخصيصها دون النظـر في اللغـات المقيـدة، فـإن جميـع العيـوب المرتبطـة بالتكـشيف والاسترجاع، باستخدام اللغـة الطبيعيـة (راجـع الفـصل الـسابع) نجـدها أيـضا في التكشيف الترابطي. وفضلا عن ذلك فإنه إذا ما كان التكشيف الترابطي يقتـصر عـلى الكلمات المفردة دون سواها، فإن ذلك يبدو مجرد قيد يوضع على الطـريقتين الهجائيـة والهرميــة القــديمتين، ولــيس طريقــة ثالثــة جديــدة للتكــشيف والاســترجاع *(Gull, 1987)*.

(*) أي أخطاء الاسترجاع. (المترجم)

ومع ذلك فإنه إسهام تاوبه في المجال لا يضارع؛ إذ إنه ابتكر التكشيف الترابطي، وأدخل المنطق البوليني في التكشيف والاسترجاع. وكما ذكر جال *Gull* *(1956)*، إيجازا فإن نظام المصطلح الأحادي أبرز الخصائص التالية، بالمقارنة بالتصنيف والفهرسة الموضوعية:

- التكلفة الأقل
- الحجم الأصغر
- التحليل الأسرع
- كفالة نقاط وصول أكثر لكل وحدة مكشفة
- معدل التعطل الأبطأ
- التخصص الأكثر
- الشمول
- البنيان المنطقي
- الحياد
- البساطة
- الملاءمة للنشر التراكمي

وليس هناك أدنى شك فعلا، في أن الدافع الحقيقي للطرق الحديثة لاسترجاع المعلومات، يرجع إلى مورتيمر تاوبه منشئ مؤسسة الوثاقة *Documentation Inc.* *(Lancaster, 1968)*. وعندما ننتقل من العصر الإلكتروني إلى العصر الرقمي، يظل تأثير تاوبه على التكشيف والاسترجاع واضحا بجلاء.

هانز بيتر لون (١٨٩٦ – ١٩٦٤):

إذا كان تاوبه يُنظر إليه بوصفه الرائد الذي وضع الأساس لاستخدام الحاسبات في استرجاع المعلومات، بإدخال المنطق البوليني في نظام المصطلح الأحادي، فإن هانز بيتر لون *Hans Peter Luhn* هو الذي ابتكر فعلا التطبيقات المعتمدة على الحاسب بالنسبة للمجال.

فقد كان لون، الذي ولد في ألمانيا، مهندسا بالمؤهل، وأصبح مخترعا شهيرا في آي بي إم IBM؛ إذ سجل أكثر من ثمانين براءة اختراع. وقد جاءت بداية لون في علم المعلومات بوجه عام، واسترجاع المعلومات على وجه الخصوص، في فترة عامي ١٩٤٧ و١٩٤٨، عندما اتصل به كل من جيمس بري *James Perry* ومالكوم دايصون *Malcolm Dyson*، وسألاه ما إذا كان من الممكن تصميم آلة آي بي إم *IBM*، للبحث في المركبات الكيميائية المشفرة *Coded*، وفقا لنظام دايصون للترقيم *(Harvey, 1978)*. وسرعان ما أبدى (لون) اهتماما بالموضوع، وانضم إليهما لتطوير واختبار نظام إلكتروني رائد للبحث عن المعلومات، ذلك الذي أسماه لون عام ١٩٤٨، منتقي البحث الإلكتروني *electronic searching selector*، وأصبح هذا الجهاز يعرف فيما بعد باسم ماسح (لون) الضوئي *Luhn scanner (Schultz, 1968)*. وفي عام ١٩٥٣ كان لون ينفق جانبا متزايدا من الوقت في استرجاع المعلومات، وقد نشر أول بحث له في المجال بعنوان "طريقة جديدة للتسجيل والبحث عن المعلومات *A New Method of Recording and Searching Information*" *(Luhn, 1953)*. كذلك أصبح لون مديرًا لبحوث استرجاع المعلومات بشركة آي بي إم. إذ استكشف وصمم كثيرا من تطبيقات استرجاع المعلومات المعتمدة على الحاسبات، التي تبدو اليوم مألوفة في المجال.

ومن بين هذه التطبيقات نظام الكلمات المفتاحية في السياق *(KWIC)* الذي ينطوي على ثلاثة عناصر أساس لتنظيم المعلومات واسترجاعها، فالعنصر الأول هو أن الكلمات المفتاحية، لا المصطلحات المستقاة من خطط التصنيف ورؤوس

الموضوعات التقليدية، تستخدم للتعبير عن جميع أوجه الوثـائق واسـترجاعها. ومـن الممكن النظر إلى الكلمات المفتاحية، على نحو مـا، بوصـفها مـن أنجـال مـصطلحات تاوبه الأحادية، على الرغم من قلة مـن حـاولوا إجـراء هـذا الـربط. وقـد اتبـع نهـج الكلمات المفتاحية هذا منذ ذلك الوقت على نطاق واسع، في تطبيقات كالتكشيف الآلي والاستخلاص الآلي، والبحث بالكلمات المفتاحية. أما العنصر الثاني لطريقة الكلـمات المفتاحية فمستمد من كشافات النصوص *concordances*، التي ابتكرهـا أسـلافنا في القرن الثالث عشر (*Wellisch, 1995*). وبينما تشكل جميع الجمل في وثيقة مـا كـشافا للنص، تشكل العناوين وما شابهها مـن عنـاصر وظيفيـة (كالجمـل التـي تـدل عـلى موضوعات الفقـرات *topic sentences* أو العنـاوين الفرعيـة) مـا يـسمى الـسياق بالنسبة لكشافات الكلمات المفتاحية في العنوان (*Key Word in Title (KWLT*، أحد امتدادات تطبيق الكلمات المفتاحية في الـسياق *KWIC*. أمـا العنـصر الثالـث في نهـج الكلمات المفتاحية في الـسياق، فهـو تباديـل الكلـمات المفتاحيـة الـواردة في العنـاوين والمقابلات الأخرى. وعـادة مـا تتخـذ التباديـل شـكلين للعـرض؛ أولهـما الكلـمات المفتاحية في السياق؛ إذ ترد كلمة مفتاحية بعينها في وسـط الـسياق، وكلمـة مفتاحيـة خـارج الـسياق (*Key Word Out of Context (KWOC*، حيـث تعـرض كلمـة مفتاحية بعينها وترد خارج السياق. والكلمات المفتاحية خارج السياق، شأنها في ذلك شأن الكلمات المفتاحية في العنوان، أحد أشكال تطبيق الكلمات المفتاحيـة في الـسياق. وقد أنتج لون الذي سك مصطلح الكلمات المفتاحية في السياق *KWIC*، عام ١٩٥٨،

بنجاح كشافا للكلمات المفتاحية في السياق للعناوين الكيميائية(*) *Chemical Titles*، ليصل بفكرته إلى نتيجة عملية *(Fischer, 1966)*. ونهج الكلمات المفتاحية في السياق معلم واضح ولا شك، في تنظيم المعلومات واسترجاعها.

ويمثل التكشيف الآلي والاستخلاص الآلي إسهامات جوهرية، قدمها لون لمجال استرجاع المعلومات؛ فباستخدام الطرق الإحصائية، طور لون وروج خوارزميات لإنتاج الكشافات والمستخلصات آليا. وفيما يتعلق بالتكشيف الآلي، تستند الإجراءات في المقام الأول، إلى التقاط الكلمات المهمة (أي الكلمات المفتاحية) التي تحمل دلالات، من الوثائق. أما الكلمات التي ترد بكثافة (أي الكلمات عالية التردد كأدوات التنكير والتعريف، وحروف العطف، وحروف الجر(**)) أو الكلمات التي نادرا ما ترد بالوثائق (أي الكلمات منخفضة التردد، أو الكلمات التي نادرا ما يستعملها البشر فى الاتصال) والكلمات التي لا تحمل دلالة أو معلومات *noninforming words* (كالأسماء العامة كتقرير وملخص، وكذلك المصطلحات التي تستعمل دائما في صيغة أو عبارة نحوية بعينها، مثل "استرجاع المعلومات"(***) في مرصد بيانات للوثائق، لأغراض استرجاع المعلومات) فيمكن استبعادها باستخدام قائمة للكلمات المستبعدة *stop-word list*، أو بأحد أساليب إحصاء تردد الكلمات *(Luhn, 1958)*. وقد أصبح هذا النهج في إنتاج كشافات الكلمات المفتاحية في

(*) خدمة إحاطة جارية في مجال الكيمياء. (المترجم)

(**) تسمى الكلمات الوظيفية أو النحوية، إذ تؤدي وظيفة نحوية ولا تحمل دلالة موضوعية. (المترجم)

(***) كمضاف ومضاف إليه، أو صفة وموصوف. (المترجم)

السياق، وتنويعاتها كالكلمات المفتاحية خارج السياق، والكلمات المفتاحية في العنوان، وكشافات الكلمات المفتاحية الأخرى ، وحتى يومنا هذا، هي طريقة التكشيف الآلي الكامل الوحيدة. وفيما يتعلق بالاستخلاص الآلي، اقترحت طريقتان للتحقق من الكلمات المهمة، ثم الجمل المهمة، التي يمكن أن تكون أكثر قدرة على تمثيل وثيقة معينة. ويشكل نهج الكلمات المفتاحية، كما سبقت الإشارة إليه، في سياق التكشيف الآلي، إحدى طريقتين لبناء المستخلصات الآلية. وتعتمد الطريقة الأخرى على الموقع النسبي للكلمات المهمة في الجملة. وبناء على ما ذهب إليه لون *Luhn (1958)*، فإن تقارب أربع أو خمس من الكلمات غير المهمة بين الكلمات المهمة، يبدو أمرا مفيدا بالنسبة لالتقاط الجمل المهمة من وثيقة ما. ويبدو الجمع بين تردد الكلمات المفتاحية، وتقارب الكلمات المفتاحية في جملة ما، طريقة مناسبة لتوليد المستخلصات الآلية.

ولقد كان إسهام لون البارز الثالث في مجال تنظيم المعلومات واسترجاعها، هو تطوير نظم البث الانتقائي للمعلومات (إس دي آي *SDI*)؛ فالبث الانتقائي للمعلومات أسلوب للبث الفعال للمعلومات العلمية الجديدة، إلى المستفيدين المستهدفين، بناء على سماتهم *profiles*(*). وقد بين لون *Luhn (1961)* معالم مقومات البث الانتقائي للمعلومات، ومختلف خطوات إدارة نظام هذا البث، ومن بين هذه المقومات يبدو وضع وصيانة سمات المستفيدين أهم المهام وأكثرها حيوية. وتشتمل سمات اهتمامات المستفيد على قائمة بالكلمات بالإضافة إلى أوزانها الراهنة، وتخضع

(*) بيان اهتماماتهم الموضوعية، والخصائص اللغوية والنوعية لما يفضلون من وثائق. (المترجم)

سمات المستفيد للمراجعة والتعديل تبعا لما يطرأ على اهتمامات المستفيد واحتياجاته من تغير. ثم تضاهى السمات بعد ذلك مقابل ما يمثل الوثائق (كالمستخلصات أو المصطلحات الكشفية)، وذلك على فترات زمنية محددة (أسبوعيا أو شهريا على سبيل المثال). وكما تصور لون فإن بالإمكان تصميم نظم ذكية، لإدارة الأعمال، والنشاط العلمي، والأنواع الأخرى من الإنتاج الفكري، وذلك بتطبيق مفهوم البث الانتقائي للمعلومات *(Stevens, 1968)*.

ومما لا شك فيه، أن لون قد وضع أفكارا قليلة حقا في حيز التطبيق لتنظيم المعلومات واسترجاعها، بمساعدة الحاسبات، حتى وإن لم تكن كل هذه الأفكار من ابتكاره *(Wellisch, 1995; Stevens, 1968)*. ولما كانت جميع التطبيقات التي سبق ذكرها تعتمد على الحاسبات، فإن كفاءة تلك العمليات المتصلة باسترجاع المعلومات قد تطورت على نحو جوهري. إلا أن هذه التطبيقات لا يمكن أن تبلغ مستوى الجودة التي تتمتع بها بوجه عام الكشافات والمستخلصات ومهام الاسترجاع التي تتم يدويا.

وقد أكسبت لون إسهاماته في المجال، وخصوصا في تنظيم المعلومات واسترجاعها، مكانة مرموقة في تاريخ علم المعلومات، الذي يشكل فيه استرجاع المعلومات مكونا رئيسا. واعتمادا على تحليل أعده كارلوس كوادرا *Carlos Cuadra (1964)*، تصدر اسم لون الجميع في ثلاث من أربع قوائم للإسهامات الجوهرية، في نظر أربعة خبراء في علم المعلومات. كذلك احتل لون المرتبة الرابعة بين خمسة وعشرين مؤلفا يتربعون على القمة من حيث غزارة النشر. وهذا رصيد نقاط قام كوادرا بحسابه اعتمادا

على البيانات الوراقية (الببليوجرافية). وفضلا عن ذلك احتل لون مكانا ضمن العشرة المتربعين على القمة في القوائم الأربع لأعلى المؤلفين المستشهد بأعمالهم بكثافة في المجال. وتؤكد جميع النتائج أن لون يبرز كباحث مرموق في علم المعلومات، وخصوصا في استرجاع المعلومات، بصرف النظر عن طريقة التقييم المستخدمة (كمشورة الخبراء، وتحليل الكتب الدراسية، وتحليل الاستشهادات المرجعية) *(Cuadra, 1964)*. وإلى لون يرجع فضل إدخال الحاسبات في مجالنا، كما كانت له الريادة في كثير من تطبيقات تنظيم المعلومات واسترجاعها، فضلاً عن كونه عاملا محفزاً للبحوث التجريبية في هذا المجال.

كالفن مورز (١٩١٩-١٩٩٤):

بالمقارنة بتاوبه ولون، جاءت إسهامات كالفن إن. مورز *Calvin N. Mooers* في تنظيم المعلومات واسترجاعها، متأخرة كثيرا. فقد كانت مجالات مورز الدراسية هي الرياضيات والفيزياء، إلا أنه كرس قدرا كبيرا من وقته للمعلومات وعلوم الحاسب، بعد التحاقه بمعهد مساتشوستس للتقنية *Massachusetts Institute of Technology*، عام ١٩٤٦، لاستثمار وتنمية خبراته في استخدام الحاسبات *(Corbit, 1992)*.

وفي عام ١٩٥٠ سك مورز مصطلح "استرجاع المعلومات" *information retrieval*، الذي ارتبط منذئذ تكامليا، وبسهولة بلغة علم المعلومات(*). وكما ذهب مورز، فإن استرجاع المعلومات يعني العثور على المعلومات التي لم يستدل على مكانها

(*) لقد سك كالفن مورز مصطلح "استرجاع المعلومات" ليستعمل بديلا عن مصطلح "الوثاقة documentation" الذي لم يكن يحظى بإجماع القبول في مجتمع الناطقين بالإنجليزية. وقد فجر سك هذا المصطلح موجة الجدل الذي لا يزال مستمرا حول مفهوم المعلومات. (المترجم)

أو على وجودها من قبل *(Garfield, 1997)*. كذلك يرجع الفضل إلى مورز *Mooers* في اقتراح قانون مورز لنظم استرجاع المعلومات:

> يمكن العزوف عن استخدام نظام استرجاع المعلومات، حيثما يكون الحصول على المعلومات بالنسبة للمستفيد أكثر مشقة وإزعاجا من عدم الحصول عليها *(P. ii)*.

وقد أعيدت صياغة قانون مورز؛ إذ تنص إحدى طبعاته على أنه "لا يمكن استخدام نظام المعلومات إلا عندما يكون عدم استخدامه أكثر إزعاجا من استخدامه" *(Koenig, 1987)*. كذلك اقترح جارفيلد *Garfield (1997)* نتيجة منطقية لقانون مورز: "كلما كانت المعلومات التي يقدمها نظام الاسترجاع صالحة، ازدادت احتمالات الإفادة منه". وقصارى القول، فإن قانون مورز يدل على نحو جوهري، على أن النظم التي ترتبط باحتياجات وممارسات من يفيدون منها، أكثر جدارة بالاستشارة عن طيب خاطر *(Henderson, 1996)*.

وبالإضافة إلى سك مصطلح "استرجاع المعلومات"، ووضع قانون مورز، طور مورز نظام ترميز زاتو *Zatocoding*، لفرز عدد كبير من واصفات الوثائق، على بطاقة واحدة مثلومة *notched* خصوصًا، وذلك بتركيب *superimposing* شفرات واصفات عشوائية من ثمانية أعداد *eight-digit descriptor codes* على هذه البطاقات. وكان من الممكن لاستخدام نظام ترميز زاتو أن يسفر عن عدد صغير فقط، يمكن التغاضي عنه من الأخطاء *false drops* في عملية البحث الوراقي *(Garfield,*

1997). كذلك كان مورز مسئولا عن ابتكار تطبيقين موجهين نحو علوم الحاسب، وهما حساب النصوص وتجميعها (تراك *Text Reckoning and Compiling TRAC*)، ولغة الحاسب ڤى إكس إم *VXM*. وكان تراك مصمما على وجه التحديد، لمعالجة النصوص غير المنظمة *unstructured*، بطريقة تفاعلية، على عكس طريقة الدفعات. أما لغة ڤى إكس إم *VXM* فكانت تستخدم لنظم الشبكات متعددة الحاسبات *(Corbit, 1992; Henderson, 1998)*.

واعترافا بإسهاماته البارزة في مجال علم المعلومات، منح مورز جائزة الاستحقاق من الجمعية الأمريكية لعلم المعلومات *American Society for Information Science Award of Merit*، عام ١٩٧٨. وتنص شهادة منح الجائزة على أنه "قد أثر في مجال المعلومات، كما أن أفكاره المبكرة تدخل الآن في واقع ممارسات اليوم" *(Henderson, 1996)*. ولقد كان مورز فعلا أحد الرواد العظام في مجال تنظيم واسترجاع المعلومات الناشئ.

جيرارد سالتون (١٩٢٧-١٩٩٥):

يتفق كل من ينتمي إلى أوساط استرجاع المعلومات، على أن جيرارد سالتون *Gerard Salton* كان واحدا من الشخصيات المرموقة في المجال؛ فقد كان الرجل "المسئول في المقام الأول عن تأسيس مجال استرجاع المعلومات وصموده والاعتراف به..." بإنفاق ثلاثين عاما من الجزء الأخير من حياته "يرعاه بعناية، ويسانده عندما يكون المناخ المهني قاسيا أو غير مواتٍ، ويدافع عنه إلى أن يتمكن من دعم نفسه"

(Grouch, et al., 1996). وإذا ما نظرنا إلى مجال استرجاع المعلومات برمته، بوصفه بنيانا تعلوه قبة، فإنه يمكن لسالتون أن يكون القبة، بينما يمكن لزملائه، ومن يلوذون به أن يقوموا مقام الأعمدة أو العناصر الأخرى المساندة للبنيان.

ولقد كانت أداة سالتون البحثية الرئيسة، نظام معالجة النصوص واسترجاعها (سمارت *System for the manipulation and retrieval of texts (SMART*، الذي يعرف أيضا من قبيل التندر بوصفه "مسترجع سالتون الآلي السحري للنصوص *Salton's magical automatic retriever of text*"، وقد غيرت أفكار سالتون طرق معالجة النصوص الكاملة بالحاسب تغيرًا جوهريًا، وأمدت مجال استرجاع المعلومات بأسس متينة *(ACM SIGIR, 1995)*. "وتشمل إسهامات سالتون البحثية، المجال الكامل لاسترجاع المعلومات؛ كأنموذج الفضاء المتجهي *vector space model*، ووزن المصطلحات، والتلقيم المرتد الخاص بالصلاحية *relevance feedback*، والتجميع العنقودي *clustering*، والاسترجاع البوليني الممتد، وقيم تميز المصطلحات، وبناء المعاجم، وتبعية المصطلحات *term dependency*، وفهم النصوص وتنظيمها، واسترجاع الفقرات، والمعالجة الآلية للنصوص، باستخدام سمارت *SMART* بالطبع" *(Crouch, et al., 1996)*. ويمكن لكل واحد من هذه الإسهامات أن يكون ولاشك، موضوعا لمناقشات مستفيضة، كما يحظى بعضها بالمناقشة في أجزاء أخرى من هذا الكتاب. إلا أنه ليس من قبيل المبالغة القول بأن سالتون قد أدخل علوم الحاسَب، والأساليب المعاصرة إلى مجال استرجاع المعلومات.

وكانت نظم استرجاع المعلومات التي تعمل على نطاق تجاري، في ستينيات القرن العشرين، تستخدم في الأساس المنطق البوليني، وتقنيات استرجاع ما قبل سمارت الأخرى. وتطبق اليوم العشرات من النظم التجارية الشهيرة الأفكار والتقنيات التي تطورت في سمارت. فقد حصلت انديفيديوال *Individual* (وهي خدمة مقتطفات إخبارية جديدة) على سبيل المثال، على ترخيص استخدام تقنيات سمارت مباشرة. وتعد نظم أخرى مثل ندل المعلومات واسعة المدى *wide area information servers* (ويز *WAIS*)، وداوكويست *DOWQUEST*، وهي إحدى أدوات الشبكة الإخبارية لداو جونز *Dow Jones Newswire*، تقنيات مشتقة من سمارت مباشرة. وقد استغرقت نظم جديدة كثيرة سنوات من البحث، ومن بينها (ون) *WIN*، وهو نظام استرجاع قانوني، تديره شركة وست للنشر *West Publishing Company*، وانكوايري *INQUERY*، وهي أداة بحث مرموقة أخرى *(ACM SIGIR, 1995)*. وهناك أساليب استرجاع معلومات، لم يتم اختبارها إلا في سمارت، وتطبق الآن بوجه عام، حتى في أحدث فئات نظم استرجاع المعلومات، كنظم استرجاع الإنترنت.

وتبدي كل هذه النظم التأثر بمفاهيم سالتون وبحوثه في استرجاع المعلومات، كما تطبق هذه المفاهيم والبحوث في البيئات الإلكترونية والرقمية. وفضلا عن ذلك كان سالتون مؤلفا غزير الإنتاج؛ فقد نشر خمسة كتب حول استرجاع المعلومات، وأكثر من ١٥٠ مقالة بحثية في المجال، طوال حياته المهنية والوظيفية *(ACM SIGIR, 1995)*. كذلك كرس سالتون جهوده البارزة للمجال، وحصل في مقابل ذلك على جوائز عدة.

كارن سبارك جونز (١٩٣٥-٢٠٠٧):

أنهت كارن سبارك جونز *Karen Spärck Jones* دراستها في المرحلة الجامعية الأولى في التاريخ أولا، ثم في الفلسفة بعد ذلك، بجامعة كمبردج بالمملكة المتحدة *University of Cambridge, U.K.* وكانت دراستها للفلسفة هي التي قادتها لمجال استرجاع المعلومات، وعلى وجه الخصوص ما يسمى الآن معالجة اللغة الطبيعية *(NLP)*. وفي رسالتها للدكتوراه عام ١٩٦٤، بعنوان "المترادفات والتصنيف الدلالي"، استخدمت نظرية الكتل *clumps* لتجميع عناقيد المصطلحات *(Wilks, 2007)*.

ولقد قدمت سبارك جونز، بعملها في استرجاع المعلومات لأكثر من أربعة عقود، إسهامات جوهرية، في مجالات عدة، كما أثرت في كثير من المحيطين بها، من الزملاء والطلبة على السواء. ومن الناحية الزمنية، كان النهج التجريبي لاسترجاع المعلومات، أول هذه المجالات. وفي غضون السنوات النهائية لبحوث سبارك جونز للدكتوراه، كانت تجارب كرانفيلد الملحمية قد دخلت مرحلتها الثانية، وبدأت سبارك جونز تبدي اهتماما قويا، لا بالنهج التجريبي الذي سلكته دراسات كرانفيلد فحسب، وإنما بدقائق الطرق والأدوات التجريبية أيضا *(Robertson & Tait, 2008)*. ثم استخدمت سبارك جونز بعد ذلك مجموعة كرانفيلد، في إجراء تجارب لاختبار التجمعات العنقودية للمصطلحات، لأغراض التصنيف الدلالي، تلك التجمعات التي بدأت تستكشفها في بحثها للدكتوراه.

غير أن بحوث سبارك التجريبية في استرجاع المعلومات، لم تتوقف عندما انتهت تجارب كرانفيلد عام ١٩٦٧، وإنما قادت جهدا مشتركا للباحثين المقيمين في المملكة المتحدة، في سبعينيات القرن العشرين، لتنمية مجموعة تجريبية جديدة "مثالية"، بحيث يصبح من الممكن إجراء التجارب، خارج حدود مشروع تجريبي واحد كتجارب كرانفيلد.

وعلى الرغم من أن هذا الجهد لم يؤت ثماره؛ نظرا لشح الدعم المالي المتوافر في المملكة المتحدة، فإنها واصلت العمل الشاق، دون كلل في هذا المجال. وقد قامت سبارك جونز بتحرير "تجارب استرجاع المعلومات" *Information Retrieval Experiment (Spärck Jones, 1981)*، الكتاب أحادي الموضوع *monograph* الوحيد المكرس في المقام الأول، للطرق التجريبية في استرجاع المعلومات، كما كتبت أيضا بحثين في ذلك الكتاب. ومن ناحية أخرى، عندما بدأت سلسلة مؤتمرات استرجاع النصوص ترِك *Text Retrieval Conference TREC*، عام ١٩٩٢، أنجزت سبارك جونز المجموعة التجريبية المثالية، التي كانت هي وزملاؤها عاجزين عن تكوينها قبل عقدين تقريبا. ومنذ البداية، شاركت سبارك جونز، بحماس في سلسلة ترِك *TREC*، بوظائف أو مهام متعددة (كمستشار غير رسمي للمنظمين، وعضو لجنة البرنامج، وعضو مشارك بفريق البحث). والأهم من ذلك، قامت سبارك جونز بكتابة سلسلة من البحوث، تلخص وتقارن أداء مختلف الفرق المشاركة في تجارب ترِك *TREC*. كذلك سجلت تأملاتها حول الدروس التي يمكن لمجتمع استرجاع المعلومات استخلاصها من العدد الهائل من النتائج المتباينة، التي انتهى إليها المشاركون في ترِك *(Spärck Jones, 1995, 2000)*. ولم يكن هناك من أحد يبدو أكثر كفاءة من سبارك جونز، للنهوض بهذه المهمة بالغة الأهمية، نظرا لاهتمامها طوال حياتها بالنهج التجريبي في استرجاع المعلومات، والتزامها تجاه هذا النهج.

ويمثل البحث الذي نشرته سبارك جونز عام ١٩٧٢، عن التردد العكسي للوثائق *inverse document frequency idf*، مجالا آخر لبحوثها، وهو مجال الطرق الإحصائية في استرجاع المعلومات *(Spärck Jones, 1972)*. وقد شرحت في ذلك البحث، الذي لا يزال على قمة البحوث المستشهد بها في المجال، كيف يمكن لوثيقة ما

أن تكون صالحة، لا لأن الكلمات المفتاحية تتردد بكثافة بها فحسب، وإنما لأن تلك المصطلحات لا تتردد بكثافة في الوثائق الأخرى غير الصالحة أيضا. وبالارتباط بخوارزمية تردد المصطلحات *term frequency (tf)*، التي طورتها مجموعة سالتون، بجامعة كورنل *Cornell University*، أصبح الجمع بين التردد العكسي للوثائق *idf*، وتردد المصطلحات، أكثر طرق وزن البديل الاحتياطي *default* استخداما لسنوات طويلة. وقد أفضى معيار التردد العكسي للوثائق *idf*، فيما بعد، إلى وزن الصلاحية في الأنموذج الاحتمالي الذي طوره روبرتصون بالتعاون مع سبارك جونز *(Robertson & Spärck Jones, 1976)*. وعلى نحو ما، يشكل جهد سبارك جونز في التردد العكسي للوثائق، مصدر إلهام لسلسلة من الدراسات التجريبية متينة الأساس لوزن المصطلحات. وقد أثبتت الطرق الإحصائية، ممثلة في التردد العكسي للوثائق، وغيره من المقاييس ذات الطابع المماثل، أنها ليست باهظة التكلفة، كما أنها تنافسية أيضا في استرجاع المعلومات. ومن الممكن الاطلاع على أحدث شاهد على ذلك في سلسلة ترِك *(Spärck Jones, 1995)*.

وبامتداد بحوثها حول الطرق الإحصائية في استرجاع المعلومات، إلى مجال أكبر مختلف، ولكنه وثيق الصلة، بدأت سبارك جونز البحث في ثمانينيات القرن العشرين، في معالجة اللغة الطبيعية، مع التركيز، بوجه خاص على التلخيص الآلي، والرد على الاستفسارات، والاستفسار باللغة الطبيعية. وكان لديها اعتقاد جازم بأن "الكلمات قائمة في حد ذاتها فقط" *(Wilks, 2007)*، وكانت ترى من العجب أن يتم تمثيل النصوص باستعمال لغات (كرؤوس الموضوعات مثلا) بخلاف اللغة الطبيعية

(Spärck Jones, 1994). وقد أكسبتها بحوثها حول معالجة اللغة الطبيعية NLP مثل: (Spärck Jones 2005, 2007) في النهاية شهرة كبيرة في ذلك المجال على الأقل، كما في المجالات الأخرى، على الرغم من أن معالجة اللغة الطبيعية تختلف عن المجالات الأخرى بالمقارنة. ونظرا لإسهام سبارك جونز المهم في معالجة اللغة الطبيعية، نشر تكريما لها كتاب بعنوان "رسم خريطة مسار جديد؛ معالجة اللغة الطبيعية واسترجاع المعلومات" Charting a New Course: Natural Language Processing and Information Retrieval (Tait, 2005). وهذا ليس الكتاب الوحيد الذي يبرز إسهامها وتأثيرها في معالجة اللغة الطبيعية. والواقع أنها بالإضافة إلى الكثير من الكتب الأخرى التي قامت بتأليفها، اشتركت في تحرير كتاب في معالجة اللغة الطبيعية في عام ١٩٨٦ (Grosz, Spärck Jones & Webber, 1986).

وكانت بحوث سبارك جونز في السنوات التالية، تشمل استرجاع الوثائق الناطقة. وبالاشتراك مع زملائها، كتبت بحثين حصلا على جوائز، تقديرا لعملهم الابتكاري حول هذا الموضوع (Maybury, 2005). وكانت الخوارزميات المحكمة غير المقيدة للكلمات المفتاحية، وأساليب استرجاع المعلومات المعتمدة على النصوص المعدلة القائمة، من بين البحوث التي قامت سبارك جونز بإجرائها حول البيانات الصوتية، كالأحاديث ونشرات الأخبار الإذاعية.

لقد أفاد مجال استرجاع المعلومات كثيرا من جهود سبارك جونز، في طريقة التردد العكسي للوثائق في وزن الوثائق، ومعالجة اللغة الطبيعية، واسترجاع الوثائق

الناطقة، والمنهج التجريبي في استرجاع المعلومـات *(Willet & Robertson, 2007)*. وسوف يظل تأثيرها في مجتمع اسـترجاع المعلومـات، كباحـث، ومعلـم، ومستـشار، ملموسا لسنوات طويلة(*).

نظرة تفصيلية في المفاهيم الأساس:

يتضمن عنوان هذا الكتـاب أربعـة مفـاهيم رئيـسة؛ هـي المعلومـات وتنظـيم المعلومات، واسترجاع المعلومات، والعصر الرقمي. ولكل هذه المفاهيم بـلا اسـتثناء مرادفات، ومن الممكن تأويلهـا وفهمهـا عـلى أكثـر مـن نحـو واحـد، في الـسياقات المختلفة. وقد قصد بالمناقشة التالية شرح معـاني ومـضامين المفـاهيم الـواردة في هـذا الكتاب وبيانها.

المعلومات:

لقد حظيت المعلومات كمفهوم بـالنظر، وقورنـت بمـصطلحات كالبيانـات، والمعرفة، والحكمة، بكثافة (كما في *Meadow, 1992* على سبيل المثال)، ومن ثم فإنه لا تبدو هناك حاجة بالنـسبة لهـذا الكتـاب، لأن يكـرر أو يـستطرد في تلـك المناقـشات والمنـاظرات. ومـن ناحيـة أخـرى، فـإن الكلـمات "المعلومـات"، و"النـصوص"، و"الوثـائق"غالبـاً مـا تـستعمل تبادليـاً في المجـال (مثـل "اسـترجاع النـصوص"، و"تنظيم الوثائق"). ووفقا لما ذهب إليه لارسن *Larsen (1999)*:

(*) نضيف هنا أن كارن سبارك جونز شاركت أيضا في تأليف كتاب حول علاقة علم اللغة بعلم المعلومات: Karen Spärck-Jones & Martin Kay. Linguistics and information science. London, Academic Press, 1973. (المترجم)

فإن الوثائق، في مجالنا، تتسم بأن لها ''سعرا''، ويمكن أن تحصى ''عددا''، وهي على هذا النحو المكونات الأساس لإحصاءات المكتبات. كما أن معظمها ''يشغل حيزا''، كما يمكن أن ''تدمر نتيجة للتداول'' كما أنها يمكن أن تتحلل أو تضمحل أو تبلى بمضي الوقت *(P. 1020)*.

وبالإضافة إلى تعريف لارسن، يمكن للوثائق أن تشمل الوسائط المتعددة. وإذا كانت النصوص تعني المعلومات النصية دون سواها، فإن ''الوثائق'' يمكن أن تشمل معلومات الوسائط المتعددة (بمعنى أي تجميع من المعلومات الصوتية والبصرية والصور، فضلا عن المعلومات النصية). ويبدو أن ''المعلومات'' تشمل كلا من النصوص والوثائق؛ أي إن مضمونها هو الأوسع بين المفاهيم الثلاثة.

وقد أجريت في السنوات الأخيرة بحوث حول استرجاع الفقرات، في مقابل استرجاع الوثائق *(Spärck Jones, 2000)*. ويدل ''استرجاع الفقرات'' (الذي يسمى أيضا استرجاع المعلومات) على العثور على المعلومات، أو الفقرات من الوثائق على وجه التحديد (كفقرات أو قطاعات من الوثائق بأطوال متفاوتة) التي يحتاج المستفيد إليها. أما ''استرجاع الوثائق''، في مقابل ذلك، فيعني الحصول على الوثيقة كاملة للمستفيد النهائي، حتى وإن كانت حاجته تقتصر على فقرة موجزة. وإذا ما نظرنا إلى كلمة ''المعلومات'' بوصفها مرادفا ''للفقرات''، كما في ''استرجاع الفقرات'' فإنه ينبغي أن يكون هناك استثناء، بالنسبة للمناقشة السابقة، فيما يتعلق بمضامين ''المعلومات''.

تنظيم المعلومات:

أيا كان الشكل الذي يمكن أن يتخذه عنصر المعلومات، فإنه بحاجة لأن يتم تنظيمه لكي يكون من الممكن استرجاعه. ويشمل تنظيم المعلومات اقتطاف بعض العناصر (كالكلمات المفتاحية أو العبارات على سبيل المثال) من الوثائق، أو تعيين مصطلحات (كالواصفات أو رؤوس الموضوعات على سبيل المثال) للوثائق، بحيث يكون من الممكن تحديد معالم مضمونها، أي تمثيلها. وعادة ما يكون من الممكن تنظيم المعلومات عن طريق أي تجمع من الوسائل التالية: الاستخلاص، والتكشيف، والتصنيف، والتلخيص، والاقتطاف. وعلى الرغم من اختلاف معاني "معالجة المعلومات" و"إدارة المعلومات" فإنه غالبا ما ينظر إليهما بوصفهما مرادفين "لتنظيم المعلومات". وبينما يدل مصطلح "معالجة المعلومات" على كيفية معاملة المعلومات لأغراض الاسترجاع، فإن مصطلح "إدارة المعلومات" يتعلق بكامل الأنشطة المرتبطة بالمعلومات، من التقاط المعلومات حتى تنظيم المعلومات.

ويستعمل مصطلح "تنظيم المعلومات" في هذا الكتاب، لتغطية مختلف الجوانب والطرق الخاصة بإعداد البدائل *surrogates* أو التمثيلات (كالكشافات والمستخلصات مثلا) لأغراض استرجاع المعلومات.

استرجاع المعلومات:

لقد حظي استرجاع المعلومات بالمعالجة، بطريقة أو أخرى، بوصفه مجالا موضوعيا، يغطي كلا من جانبي تنظيم المعلومات واسترجاعها *(Spärck Jones &*

Willett, 1997). كذلك يشار إلى الجانب الخاص بالاسترجاع بوصفه أيضا "الوصول إلى المعلومات *information access*" و"التماس المعلومات *information seeking*" و"البحث عن المعلومات *information searching*". ومن الممكن النظر إلى هذه المصطلحات بوصفها مرادفات "للاسترجاع"، إلا أن لكل منها توجهه الخاص فيما يتعلق بالمضامين؛ فمصطلح "الوصول إلى المعلومات" يركز على الجانب الخاص بتحصيل المعلومات أو الحصول عليها. وفى مقابل ذلك يركز "التماس المعلومات" على المستفيد المشارك على نحو إيجابي في العملية. أما فيما يتعلق "بالبحث عن المعلومات" فإن تركيز الاهتمام يبدو منصبا على كيفية البحث عن المعلومات.

وبالإضافة إلى المصطلحات التي سبق ذكرها، فإن "التنقيب عن البيانات *data mining*" و"اكتشاف المصادر *resource discovery*"، غالبا ما نجدهما في لغة اختصاصيي المعلومات، في السنوات الأخيرة، عند مناقشة استرجاع المعلومات. وعادة ما يستخدم كلا المصطلحين في إدارة الأعمال وبيئة المشابكة. ولازلنا الآن في انتظار ما إذا كان من الممكن لكل من "التنقيب عن البيانات" و"اكتشاف المصادر" أن يتم استيعابهما في اللغة المألوفة لمن ينتمون إلى مجال استرجاع المعلومات.

وهناك طبقة دلالية أخرى في استرجاع المعلومات، وهي "اختزان المعلومات *information storage*"، الذي يتناول في المقام الأول، تسجيل المعلومات واختزانها، إلا أن هذا الاستعمال في سبيله لأن يصبح تدريجياً من ممارسات الماضي، نظرا لأن "اختزان المعلومات" لم يعد يشكل هما ذا بال، وذلك بفضل التطورات التي

حـدثت في تقنيـات اختـزان المعلومـات والوصـول إليهـا. ولهـذا، فـإن اسـترجاع المعلومات، في هذا الكتاب، يشمل الـسعي وراء المعلومـات، أو الـتماس المعلومـات، والبحث عن المعلومات، والوصول إلى المعلومات، إلا أنه يستبعد اختزان المعلومات.

العصر الرقمي:

لكلمة "رقمي *digital*" مقارنة بكلمة "تنـاظري *analog*"، مفهـوم حـديث نسبيا؛ فكل من "رقمي" و"تناظري" مصطلحان يتـصلان بالتقنيـات الإلكترونيـة، وفقا للتقرير التالي الذي قدمتـه تـك تارجـت *Tech Target (2001)*، وهـي شركـة تعرِّف أيضا بالمصطلحات في تقنيات المعلومات:

> يدل رقمي على التقنيات الإلكترونيـة التـي تولِّـد وتختـزن وتعالج أو تجهز البيانات في حالتين: إيجابيـة وغـير إيجابيـة. ويـتم التعبير عن الإيجابية أو تمثيلها بالعدد (١)، بينما يتم التعبير عن غير الإيجابية أو تمثيلها بالعدد (صفر). ومن ثم، فإن البيانات التي يـتم نقلها أو اختزانها بالتقنيات الرقمية، يتم التعبير عنها كسلسلة مـن الأصفار والآحاد ... وفـيما قبـل التقنيـات الرقميـة، كـان النقـل الإلكتروني يقتصر على التقنيات التناظرية، التي تنقل البيانـات في شكل إشارات إلكترونية متفاوتـة الـتردد أو النطـاق، تـضاف إلى موجـات حامـل تـردد معـين. وكانـت الإذاعـة والنقـل الهـاتفي يستعملان تقليديا تقنية تناظرية.

ومع ظهور الحاسبات والإنترنت وتقنيات المعلومات الأخرى، بدأنا ندخل بجلاء، العصر الرقمي. وهناك تزايد في مختلف الأنشطة، التي تشمل البحث والتطوير، المتصلة بتنظيم المعلومات واسترجاعها، التي تتم في البيئة الرقمية. ومن ثم فإن هذا الكتاب الذي صيغ عنوانه بناء على ذلك، يعني الوضع الراهن للمعرفة بالنسبة للمجال.

المكونات الرئيسة:

من الممكن تقسيم تنظيم المعلومات واسترجاعها برمته، إلى عدة عناصر رئيسة: وهي مراصد البيانات، وتقنيات البحث، واللغة، وواجهة التعامل. والبشر (بمن فيهم المستفيد، واختصاصي المعلومات، ومسئول تطوير النظم)، والمعلومات، والنظم، عناصر ثلاثة مترابطة، تعمل معا في قطاع تنظيم المعلومات واسترجاعها، على الرغم من أنها لا تحظى بالمناقشة على وجه التحديد في هذا القسم.

مراصد البيانات:

تشكل مراصد البيانات أو قواعد البيانات *databases*، في تنظيم المعلومات واسترجاعها، المعلومات التي يتم التعبير عنها أو تمثيلها وتنظيمها بطريقة معينة. وبالمعنى التقليدي يتكون مرصد البيانات (كمرصد البيانات على الخط المباشر على سبيل المثال) عادة من تسجيلات *records*، يتم تقسيمها إلى حقول *fields*، وهي أصغر الوحدات وأقربها للطبيعة، بالنسبة لفرز المعلومات والبحث عنها

واسترجاعها؛ ففي مرصد بيانات مقالات الدوريات، على سبيل المثال، يشكل المؤلف والعنوان حقلين. وتتكون مراصد البيانات التقليدية من قسمين؛ "الملفات التسلسلية *sequential files*" و"الملفات المصنفة أو المقلوبة *inverted files*". والملف التسلسلي هو مصدر مرصد البيانات؛ إذ يشتمل على المعلومات منظمة في بنيان المرصد، القائم على الحقل والتسجيلة. ويسمى بالملف التسلسلي لأن ما يشتمل عليه من تسجيلات، ترد مرتبة وفقا للتسلسل الذي تتخذه عند إدخالها في مرصد البيانات. أما الملف المصنف أو المقلوب، ويعرف أيضا بالكشاف، فيكفل الوصول إلى الملف التسلسلي، بناء على استفسارات بحث معينة. ويسمى بالملف المقلوب نظراً لأن الترتيب الذي يتم به تمثيل المعلومات (نقطة الوصول *access point* أولا ومحدد المكان ثانيا) عكس الترتيب في الملف التسلسلي (حيث محدد المكان *locator* أولا ونقطة الوصول ثانيا).

وبالمعنى غير التقليدي، يمكن لمرصد البيانات (كما في نظم استرجاع الإنترنت على سبيل المثال) أن تكون به ملفات تسلسلية وملفات مقلوبة. إلا أن تكوين الملف التسلسلي لنظام استرجاع الإنترنت، على سبيل المثال، يختلف عما يناظره في النظم التقليدية لاسترجاع المعلومات (كنظم الخط المباشر مثلا)؛ إذ لا يتخذ تكوين الملف التسلسلي شكل مرصد البيانات القائم على الحقل والتسجيلة؛ حيث يتكون هذا الملف من معلومات بلا حقول، يتم التعبير عنها بصيغ أشبه بالنثر *proselike*. يضاف إلى ذلك أن المعلومات المتضمنة في هذا الملف، ليست بدائل (كالمستخلصات مثلا) وإنما أجزاء من المحتوى الكامل لإحدى وثائق الإنترنت الأصلية (كصفحة العنكبوتية

webpage على سبيل المثال). إلا أنه في النظم التقليدية لاسترجاع المعلومات، عادة ما تكون المعلومات في الملفات التسلسلية ضربا من التمثيل في شكل توصيفات وراقية، ومستخلصات، وملخصات، ومقتطفات، وما شابه ذلك.

ويحدد محتوى مرصد البيانات وتغطيته، ما يمكن أن يتم استرجاعه فيما بعد، من نظام استرجاع المعلومات.

تقنية البحث:

لا يمكن البحث عن المعلومات التي يتم تمثيلها وتنظيمها نسقيا في مرصد البيانات، واسترجاعها إلا عندما تتوافر تقنية بحث مناظرة. ويمكن لتقنية البحث أن تكون على أي درجة من التطور، من حيث قدرات البحث، التي تتحدد معالمها في النهاية، كخوارزميات البحث، والإجراءات التي ينطوي عليها نظام استرجاع المعلومات. ومن الممكن لجميع إجراءات البحث، أن تصنف بوصفها أساسية أو متطورة. وعادة ما نجد إجراءات البحث الأساسية في معظم نظم استرجاع المعلومات العاملة، بينما يتم اختبار إجراءات البحث المتطورة، وإجراء التجارب عليها في المختبرات، أو نماذج التنظيم الأولية *prototype* في المقام الأول. إلا أنه يتزايد في السنوات الأخيرة ربط خوارزميات البحث المتطورة تكامليا بنظم استرجاع الإنترنت.

وتنتمي الإجراءات، كالبحث بالكلمات المفتاحية، والبحث البوليني، والبتر *truncation*، والبحث بالتقارب *proximity*، لمنظومة خوارزميات البحث الأساسية. وكما سبق أن أشرنا، فإنه غالبا ما تشكل هذه الإجراءات تقنيات البحث

الكامنة في كثير من نظم استرجاع المعلومات. ومن الممكن للمستفيدين النهائيين من محدودي التدريب وخبرات البحث، أن يكونوا قادرين على إنجاز مهام البحث البسيطة، في بيئات الاسترجاع من هذا الضرب. وتستخدم تقنيات البحث المتطورة، كالبحث بالوزن، في غالب الأحيان، في نظم الاسترجاع الأحدث، التي يتم تصميمها بوجه عام لأولئك الحاصلين على تدريب مهني، المتمرسين في البحث.

وتحدد قدرات تقنية البحث، ما يمكن أنه يتاح من أساليب الاسترجاع للمستفيدين، وكيف يمكن استرجاع المعلومات المختزنة في مراصد البيانات.

اللغة:

تعتمد المعلومات على اللغة، المنطوقة أو المكتوبة، عندما يتم تجهيزها، أو نقلها أو التواصل بها. واللغة في هذا السياق أحد المكونات الجوهرية في تنظيم المعلومات واسترجاعها. ومن الممكن للغة في تنظيم المعلومات واسترجاعها، أن تكون لغة طبيعية أو لغة مقننة *controlled vocabulary*. ويسمى ما يستعمله البشر على نحو طبيعي للتعبير عن المعلومات أو صياغة الاستفسارات، "لغة طبيعية *natural language*". وإذا ما استخدمت لغة اصطناعية، محدودة المفردات، والنظم *syntax*، والدلالة، والسياقات، في التعبير عن المعلومات واسترجاعها، فإن تلك اللغة تسمى "اللغة المقيدة *controlled vocabulary*" *(Wellisch, 1995)*. وهناك ثلاثة أشكال مألوفة للغة المقيدة، وهي خطط التصنيف، وقوائم رؤوس الموضوعات، والمكانز *thesauri*، ولكل منها وجه الاستخدام الخاص بها، في تنظيم المعلومات واسترجاعها.

وتكفل اللغة الطبيعية، بوجه عام، أعلى درجات التخصيص والمرونة في تنظيم المعلومات واسترجاعها. ولا يحتاج الإنسان إلى أي تدريب أو تمرين في استخدام اللغة الطبيعية؛ لأنها هي اللغة التي نستخدمها للتواصل الشفوي أو التحريري يوميا. وفي مقابل ذلك، فإن ابتكار اللغة المقيدة وصيانتها أمر باهظ التكلفة، كما أنه يتعين على المرء أن يتعلم كيف يستخدم مثل هذه اللغة بالممارسة والتدريب. إلا أنه بإمكان اللغة المقيدة الحد من المشكلات الجوهرية *intrinsic* المرتبطة باللغة (كالتعقد، والفروق الدقيقة *subtleness*، والغموض) عند استعمال اللغة الطبيعية لأغراض تنظيم المعلومات واسترجاعها *(Lansdale & Ormerod, 1994)*. ولقد كان الجدل في هذا المجال حول اللغة الطبيعية في مقابل اللغة المقيدة، محتدما منذ نهاية القرن العشرين *(Rowley, 1994)* إلا أن اللغة تظل عنصرا جوهريا في تنظيم المعلومات واسترجاعها، أيا كان ما ينتهي إليه الجدل.

فاللغة، إلى حد ما، تحدد مدى المرونة والاصطناعية في تنظيم المعلومات واسترجاعها. والفصل الرابع في هذا الكتاب مكرس لمناقشة اللغة في تنظيم المعلومات واسترجاعها.

واجهة التعامل:

واجهة التعامل *interface*، كما ذهب شو *Shaw (1991)*، هي ما يراه المستفيد، ويسمعه، ويلمسه عند التعامل مع نظام الحاسب. وتدل واجهة التعامل، في تنظيم المعلومات واسترجاعها، على التفاعل الذي يتم بين المستفيد والأنشطة المتصلة به. ومع هذا المكون في تنظيم المعلومات واسترجاعها أيضا، يبدو ما يتصل بالمستفيد

واضحا، وإن كان متداخلا مع المكونات الثلاثة الأخرى، وهي مرصد البيانات، وتقنية البحث، واللغة.

ودائما ما ينظر في واجهة التعامل عند الحكم ما إذا كان نظام معالجة المعلومات واسترجاعها متعاطفا مغ المستفيد *user-friendly*. وكما تم تعريفها في قانون مورز، فإن النظم المتعاطفة مع المستفيد، من شأنها أن تجتذب أعدادا من البشر أكثر مما تجتذب النظم المعادية للمستفيد *user-hostile* من حيث الاستخدام. وتتحدد جودة واجهة التعامل بطريقة التفاعل (كالاختيار من قائمة الخيارات *menu* فعلا) وخصائص العرض (كإخراج الشاشة وشكل الحرف مثلا) والعوامل المتصلة الأخرى. وقد بدأ استغلال التقنيات التكيفية والمؤثرة، في تصميم واجهات التعامل وتنفيذها، حيث يتجه المزيد من الاهتمام نحو البعد الإنساني في تنظيم المعلومات واسترجاعها. وتتحكم واجهة التعامل في النجاح النهائي للنظام بالنسبة لتنظيم المعلومات واسترجاعها، وخصوصا إذا ما كان النظام يعمل في البيئة الرقمية.

وقصارى القول، فإن مرصد البيانات، وتقنية البحث، واللغة، وواجهة التعامل تشكل المكونات الأساس لتنظيم المعلومات واسترجاعها، على النحو الذي يتفاعل مع البعد الإنساني، في مرحلة أو أخرى، أثناء عملية تنظيم المعلومات واسترجاعها.

المشكلة الجوهرية في تنظيم المعلومات واسترجاعها:

لا تزال المشكلة الجوهرية في تنظيم المعلومات واسترجاعها، هي كيف يمكن الحصول على المعلومات المناسبة للمستفيد المناسب، في الوقت المناسب، على الرغم من وجود متغيرات أخرى (كخصائص المستفيدين أو حدود تغطية مراصد البيانات

مثلا) في بيئة تنظيم المعلومات واسترجاعها. وقبل الاستطراد في استكشاف المشكلة الجوهرية، ينبغي النظر أولا في عملية تنظيم المعلومات واسترجاعها.

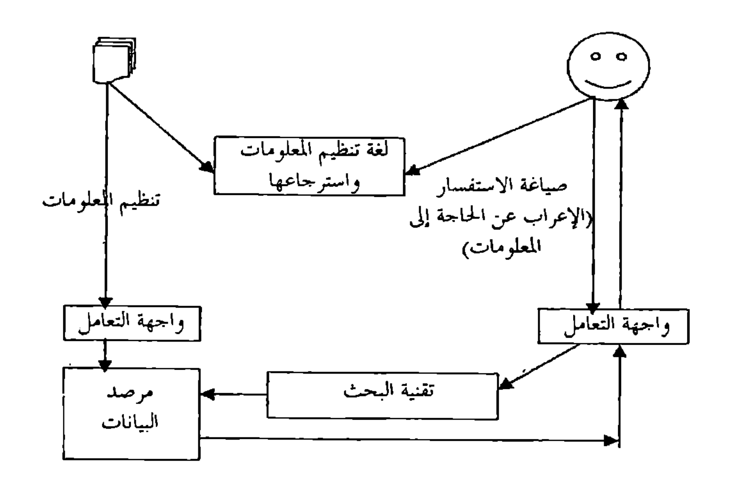
عملية تنظيم المعلومات واسترجاعها:

في عملية تنظيم المعلومات واسترجاعها، يشرع المستفيد في عملية البحث، ويتلقى أي نتائج يتم استرجاعها، بينما يعد اختصاصي المعلومات مسئولا عن تصميم نظم استرجاع المعلومات وتنفيذها وصيانتها. ويوضح الشكل رقم ١/١ عملية تنظيم المعلومات واسترجاعها.

وينبغي لأي معلومات يتم استرجاعها من مرصد البيانات، أن يتم تمثيلها أو التعبير عنها أولا من جانب اختصاصي المعلومات، وفقا للغة التي وقع عليها الاختيار لأغراض تنظيم المعلومات واسترجاعها. ومن الممكن أن تحدث بعض مظاهر التفاوت في أثناء التعبير عن المعلومات، ومن الممكن أن تكون هناك مشكلات عويصة، إذا ما استخدمت لغة مقيدة، وذلك للأسباب التالية: أولا، عندما يتم التعبير عن المعلومات المسجلة في أوعية كمقالات الدوريات أو التقارير التقنية، في شكل مستخلصات أو مصطلحات كشفية، وما شابه ذلك، فإن استعادة المعلومات الأصلية لطبيعتها، لا تبدو قابلة للتحقيق. ويمكن القول جدلا بأننا نمثل، على سبيل المثال، دائرة كبيرة بدائرة صغيرة، فإذا حدث ذلك، فإن بعد الحجم يكون قد تم تشويهه. ثانيا، إن أي لغة مقيدة مجرد مجموعة فرعية من اللغة الطبيعية التي كتبت بها الوثائق الأصلية؛ ولهذا فإنه غالبا ما يكون من الصعب أن نجد، على سبيل المثال، تطابقا تاما

أو مضاهاة دقيقة بين مصطلح في وثيقة ما، وواصفة مستقاة من مكنز ما. ولأغـراض التمثيل أو التنظيم فإنه يمكن للمكشف حينئذ، أن يختار مـن المـصطلحات المتـصلة، تلك المصطلحات الأضيق أو الأكثـر اتـساعا، الـواردة في المكنـز. ثالثـا، يبـدو عـدم الاطراد في التعبير عن المعلومـات (بـما في ذلـك تحليـل المفـاهيم) أمـرا لا مفـر منـه، وخصوصا إذا كان هناك أكثر من شخص، أو أكثر من نظام يعالج المهمة. وقد سـجل كلفـردون *Cleverdon (1984)* أنـه حتـى وإن كـان هنـاك مكـشفان متمرسـان، يستخدمان اللغة المقيدة نفسها، فإنهما يمكن أن يتفقا معا على تخصيص ٣٠ بالمئة فقـط من المصطلحات للوثيقـة نفـسها. وعـلى النحـو نفـسه، تبـين لهمفـري *Humphrey (1992)*، أن الاطـراد فـيما بـين المكـشفين في المـدلاين *MEDLINE*، في اختيـار المصطلحات مـن قائمـة رؤوس الموضـوعات الطبيـة *Medical Subject Hedings (MeSH)*، كان أقل من ٤٩ بالمئة.

ومن ناحية أخـرى، يطلـب مـن المـستفيدين تحويـل احتياجـاتهم المعلوماتيـة، باسـتخدام لغـة تنظـيم المعلومـات واسـترجاعها التـي يقـع عليهـا الاختيـار، إلى استفسارات يمكن بعدئذ تنفيذها في مرصد البيانات، بتقنية البحـث المتـوافرة. ولقـد كان الباحثون منذ زمن بعيد، على دراية بالتعقد الذي يكتنف هذه المهمـة؛ فقـد أشـار بلير ومارون *Blair and Maron (1985)*، على سبيل المثال، إلى "أنه من الـصعب إلى حد الاستحالة بالنـسبة للمـستفيدين، أن يتكهنـوا بـالكلمات وتجميعـات الكلـمات، والعبارات التي تستعمل من جانب جميع (أو معظم) الوثائق" *(P. 295)*.



الشكل رقم ١/ ١ عملية تنظيم المعلومات واسترجاعها

يضاف إلى ما سبق أن أي استعمال للغات المقيدة، ومقومات البحث (كالمعاملات البولينية مثلا) من شأنه أن يضاعف الصعوبة. والبحث في اللغة الطبيعية، أي البحث بالعبارات أو الجمل الكاملة، كما تستعمل في التواصل اليومي، دون صياغة أي استفسارات مقننة كـ(لما السماء زرقاء؟ مثلا) في سبيله لأن يكون متاحا على الإنترنت، إلا أنه لا يزال هناك طريق طويل ينبغي أن يقطع، لكي يحقق الباحثون في معالجة اللغة الطبيعية، أحد أفرع الذكاء الاصطناعي (*AI*) اختراقات في جهودهم.

وبعبارة أخرى، فإن إمكان نجاح عملية بحث ما أو عدمه يتوقف فقط، على ما إذا كانت هناك مضاهاة بين المعلومات التي تم التعبير عنها في كل من النظام والاستفسار الذي يقدمه المستفيد. ولكي نكون أكثر تحديدا، فإن عملية البحث يمكن

أن تكون ناجحة، إذا كان هناك تطابق تام أو مضاهاة، بين الاستفسار والمعلومات التي تم التعبير عنها أو تمثيلها في مرصد البيانات، الذي وقع عليه الاختيار للمهمة. وما لم يكن الأمر كذلك، فإن عملية البحث لا يمكن أن تسفر عن أي نتائج مفيدة. ولهذا فإن المضاهاة هي التقنية الأساس في تنظيم المعلومات واسترجاعها. وكما يتبين من الشكل رقم ١ / ١، فإن هناك عدة نقاط في عملية تنظيم المعلومات واسترجاعها، يمكن أن تكون وراء مظاهر التفاوت أو التضارب في المضاهاة. ولا يتحقق الهدف النهائي لجودة تنظيم المعلومات واسترجاعها، إلا من خلال استعمال مختلف الطرق والأساليب، وذلك للحد قدر الإمكان أو التخلص كلية، من مظاهر التفاوت أو التضارب التي يمكن أن تحدث أثناء العملية. ونناقش هذه الطرق والأساليب تفصيلا فيما يأتى من فصول.

أوجه القصور في تنظيم المعلومات واسترجاعها:

في الوقت الذي أجريت فيه أعداد هائلة من البحوث في تنظيم المعلومات واسترجاعها، لا تزال هناك أوجه قصور معينة يبدو من الصعب تخطيها. وحينما تأمل سوانسون *Swanson (1988)* الموقف بالنسبة للتكشيف والاسترجاع الآليين، استعار مصطلح مسلمات العجز *postulates of impotence (PIS)*، من إي. تايلور هويتكر *E. Taylor Whittaker*، وصاغ تسع مسلمات عجز، تبين الأشياء التي لا يمكن أن تتم في تنظيم المعلومات واسترجاعها. وعلى الرغم من حدوث ذلك عام ١٩٨٨، فيما يتعلق بالمجال الآلي، فإن هناك عدة عبارات لا تزال تبدو مناسبة، ونقتبسها هنا:

مسلمة العجز ١: لا يمكن الإعراب عن الحاجة إلى المعلومات كاملة كطلب بحث... فالسؤال لا يمكن أن يصاغ بدقة إلا بعد العثور على الإجابة.

مسلمة العجز ٣:... الصلاحية ليست ثابتة، وإنما يتم الحكم عليها في حدود إطار متغير.

مسلمة العجز ٤: لا يمكن مطلقا التحقق مما إذا كانت جميع الوثائق الصالحة لأي طلب قد تم العثور عليها...

مسلمة العجز ٩: قصارى القول، أن المسلمات الثماني الأوليات تعني أن التكشيف والاسترجاع الآليين الكاملين الفعالين باطراد، ليسا من المستحيلات. والمشكلات المفاهيمية لاسترجاع المعلومات، أي مشكلات المعنى أو الدلالة، ليست أقل عمقا من التفكير أو أي أشكال أخرى للسلوك الذكي *(P. 96)*.

والمشكلات المفاهيمية لاسترجاع المعلومات، كما عبر عنها سوانسون *Swanson*، حاسمة بالنسبة للإلمام بالمجال وتطوره. ويتضح من إمعان النظر في عملية تنظيم المعلومات واسترجاعها، أن ما يتضمنه هذا النشاط، كما سبق أن بينا، إنما هو في الأساس مضاهاة المصطلحات، وليس البحث عن المفاهيم في البيئة الرقمية؛ فعندما يكون مصطلح عملية البحث، على سبيل المثال "النقل العام *public transportation*"، فإن الوثائق المكشفة تحت "الحافلات" أو "مترو الأنفاق" لا يمكن أن تسترجع، ما لم يتم إعداد إحالات في اللغة المقيدة. هل يمكن لتنظيم المعلومات واسترجاعها أن يتجاوز يوما ما حدود مضاهاة المصطلحات، ويقترب أكثر من البحث عن المفاهيم؟ ويتم البحث عن الإجابة واستكشافها بهمة عن طريق التجربة والخطأ *(Swanson, 1977)*.

المراجع

ACM SIGIR. (1995): Gerard Salton: In memoriam. IRList Digest, 12(34). Retrieved January 19, 2009, from www.cs.virginia.edu/-clv2m/salton.txt.

Blair, David C. and Maron, M.E. (1985): An evaluation of retrieval effectiveness for a full-text document-retrieval system. Communications of the ACM, 28(3): 289-299.

Boole, George. (1854). An investigation into laws of thought, on which are founded the mathematical theories of logic and probabilities. London: Walton and Maberley.

Borko, Harold and Bernier, Charles L. (1975). Abstracting concepts and methods. New York: Academic Press.

Bush, Vannevar. (1945). As we may think. Atlantic Monthly, 176(1), 101-108.

Cleverdon, C.W. (1984). Optimizing convenient online access to bibliographic databases. Information Services and Use, 4, 34-47.

Corbitt, Kevin D. (1992). Calvin N. Mooers papers, 1930-1978 (CBI 81). Minneapolis: Center for the History of Computing, Charles Babbage Institute, University of Minnesota. Retrieved October 3, 2009, from www.libsci.sc.edu/bob/isp/mooers.htm.

Crouch, Carolyn, et al. (1996). In memoriam: Gerard Salton, March 8, 1927-August 28, 1995. Journal of the American Society for Information Science, 47(2), 108-115.

Cuadra, Carlos A. (1964). Identifying key contributions to information science. American Documentation, 15(4), 289-295.

Fischer, M. (1966). The KWIC index concept: A retrospective view. American Documentation, 17(2), 57-70.

Garfield, Eugene. (1997). A tribute to Calvin N. Mooers, a pioneer of information retrieval. The Scientist, 11(6), 9.

Grosz, Barbara J., Spärck Jones, Karen and Webber, Bonnie Lynn (Eds.). (1986). Readings in natural language processing. Los Altos, CA: Morgan Kaufmann.

Gull, Cloyd Dake. (1956). Seven years of work on the organization of materials in the special library. American Documentation, 7(1-5), 320-329.

Gull, Cloyd Dake. (1987). Historical note: Information science and technology: From coordinate indexing to the global brain. Journal of the American Society for Information Science, 38(5), 338-366.

Hahn, Trudi Bellardo. (1996). Pioneers of the online age. Information Processing and Management, 32(1), 33-48.

Harvey, John F. (1978). Luhn, Hans Peter (1896-1964). In Bohdan S. Wynar (Ed.), Dictionary of American library biography (pp. 324-326). Littleton, CO: Libraries Unlimited.

Henderson, Madeline M. (1996). In Memoriam: Calvin N. Moores, October 24, 1919-December 1, 1994. Journal of the American Society for Information Science, 47(9), 659-661.

Humphrey, Susanne M. (1992). Indexing biomedical documents. From thesauri to knowledge-based retrieval systems. Artificial Intelligence in Medicine, 4, 343-371.

Koenig, Michael E. D. (1987). The convergence of Moore's/Mooers' Laws. Information Processing & Management, 23(6), 358-592.

Lancaster, F. Wilfrid. (1968). Information retrieval systems: Characteristics, testing and evaluation. New York: Wiley.

Lansdale, Mark W. and Ormerod, Thomas C. (1994). Understanding interfaces: A handbook of human-computer dialogue. Academic Press.

Larsen, Poul Steen. (1999). Books and bytes: Preserving documents for prosperity. Jounal of the American Society for Information Science, 50(11), 1020-1027.

Luhn, H. P. (1953). A new method of recording and searching information. American Documentation, 4(1), 14-16.

Luhn, H. P. (1958). The automatic creation of literature abstracts. IBM Journal of Research and Development, 2(2), 159-165.

Luhn, H. P. (1961). Selective dissemination of new scientific information with the aid of electronic processing equipment. American Documentation, 12(2), 131-138.

Maybury, Mark T. (2005). Karen Spärck Jones. In John I. Tait (Ed.), (2005). Charting a new course: Natural language processing and information retrieval (pp. xi-xxiii). New York: Springer.

McCandless, R. E. J., Skweir, E. A. and Gordon, M. (1964). Secondary journals in chemical and biological fields. Journal of Chemical Documentation, 4(2), 147-153.

Meadow, Charles T. (1992). Text information retrieval systems. San Diego: Academic Press.

Mooers, Calvin N. (1960). Mooer's Law: Or, why some retrieval systems are used and others are not. American Documentation, 11(3), ii.

Pao, Miranda Lee. (1989). Concepts of information retrieval. Englewood, CO: Libraries Unlimited.

Robertson, Stephen and Spärck Jones, Karen. (1976). Relevance weighting of search terms. Journal of the American Society for Information Science, 27(3), 129-146.

Robertson, Stephen, and Tait, John (2008). In memoriam: Karen Spärck Jones. Journal of the American Society for Information Science, 59(5), 852-854.

Rowley, Jennifer. (1994). The controlled versus natural indexing languages debate revisited: A perspective on information retrieval practice and research. Journal of Information Science, 20(2), 108-119.

Salton, Gerard. (1987). Historical notes: The past thirty years in information retrieval. Journal of the American Society for Information Science, 38(5), 375-380.

Schultz, Claire K. (Ed.). (1968). H. P. Luhn: Pioneer of information science: Selected works. New York: Spartan Books.

Shaw, Debora. (1991). The human-computer interface for information retrieval. Annual Review of Information Science and Technology, 26, 155-195.

Shera, Jesse H. (1978). Taube, Mortimer (1910-1965). In Bohdan S. Wynar (Ed.), Dictionary of American library biography (pp. 512-513). Littleton, CO: Libraries Unlimited.

Smith, Elizabeth S. (1993). On the shoulders of giants: From Boole to Shannon to Taube: The origins and development of computerized information from the mid-19th century to the present. Information Technology and Libraries, 12(2), 217-226.

Spärck Jones, Karen. (1972). A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation, 28(1), 11-21.

Spärck Jones, Karen (Ed.). (1981). Information retrieval experiment. London: Butterworths.

Spärck Jones, Karen. (1994). Finding the information wood in the natural language tree [Videotape]. Talk presented at the Grace Hopper Celebration of Women in Computing meeting. 41 min.

Spärck Jones, Karen. (1995). Reflection on TREC. Information Processing & Management, 31(3), 291-314.

Spärck Jones, Karen. (2000). Further reflections on TREC. Information Processing & Management, 36(1), 37-85.

Spärck Jones, Karen. (2005). Some points in a time. Computational Linguistics, 31(1), 1-14.

Spärck Jones, Karen. (2007). Automatic summarizing: The state of the art. Information Processing & Management, 43(6), 1449-1481.

Spärck Jones, Karen, and Willett, Peter (Eds.). (1997). Readings in information retrieval. San Francisco: Mogran Kaufmann.

Stevens, Mary Elizabeth. (1968). H. P. Luhn, information scientist. In Claire K. Schultz (Ed.), H.P. Luhn: Pioneer of information science: Selected works (pp. 24-30). New York: Spartan Books.

Swanson, Don R. (1977). Information retrieval as a trial-and-error process. Library Quarterly, 47(2), 128-148.

Swanson, Don R. (1988). Historical note: Information retrieval and the future of an illusion. Journal of American Society for Information Science, 49(2), 92-98.

Tait, John I. (Ed.) (2005). Charting a new course: Natural language processing and information retrieval. New York: Springer.

Tech Target. (2001). Digital. Retrieved December 3, 2008, from whatis.techtarget.com.

Wellisch, Hans H. (1995). Indexing from A to Z. 2nd ed. New York: H.W. Wilson.

Wilks, Yorick. (2007). In memoriam: Karen Spärck Jones (1935-2007). IEEE Intelligent Systems, 22(3), 8-9.

Willett, Peter, and Robertson, Stephen. (2007). In memoriam: Karen Spärck Jones. Journal of Documentation, 63(5), 605-608.

Wynar, Bohdan S. (1980). Introduction to cataloguing and classification. 6th ed. Littleton, CO: Libraries Unlimited.

الفصل الثاني

تنظيم المعلومات - الطرق الرئيسة

تحتاج المعلومات المسجلة رسميا، في مختلف أنواع الوثائق (كبحوث المؤتمرات، والتقارير التقنية على سبيل المثال) لأن يتم تنظيمها لكي يكون من الممكن استرجاعها. ولا يمكن بحال أن يتم استرجاع المعلومات بالوثائق الأصلية نفسها، وإنما تتم عملية الاسترجاع ببدائل الوثائق، في شكل كشافات، ومستخلصات، وملخصات، وما شابه ذلك. ويدل التنظيم في هذا الكتاب على الجوهر أو المحتوى الموضوعي للوثائق بواسطة طرق معينة، وذلك على الرغم من أن الناتج النهائي يمكن أن يتخذ أشكالا مختلفة. وينبغي في الظروف المثالية، أن تتم عملية التنظيم ببساطة وكفاءة، إلا أنه كما لاحظ لسك **Lesk (1997)**:

> إذا كانت لدينا خطة واحدة لتنظيم المعرفة، تكفل لنا وضع كل فكرة في مكان واحد، وإذا كان المستفيدون على دراية بهذه الخطة، وبإمكانهم وضع أي من استفساراتهم في إطارها، فإن الاسترجاع الموضوعي يمكن أن يكون واضح المعالم... ولا يبدو عمليا، من الممكن لأي خطة لتنظيم المعرفة، أن تخدم جميع الأغراض **(pp. 99-100)**.

وفضلا عن ذلك، فإن الطريقة التي يمكن أن تستخدم بها خطة التنظيم، على نحو مطرد ودقيق، تثير تحديات لاختصاصيي المعلومات، حتى وإن كان اختيار الخطط لا يشكل هما. كما أن بعض الطرق (كالاستخلاص مثلا) لا تتطلب خطة لكي يتم تنفيذها، فتنظيم المعلومات إطار يعني التحدي والتعقد والفروق الدقيقة.

التكشيف:

لقد كان التكشيف من الطرق التي تستخدم على نطاق واسع لتنظيم المعلومات. ويستخدم التكشيف المصطلحات (أي الكلمات أو العبارات) سواء كانت هذه المصطلحات مستقاة أو مشتقة من النص، أو معينة للنص، وذلك للتعبير عن الجوانب المهمة للوثيقة الأصلية. ويتراوح عدد المصطلحات الكشفية، التي تستعمل للتعبير عن محتوى وثيقة ما، ما بين بضعة مصطلحات في مراصد البيانات الوراقية، وعدة مئات في نظم النصوص الكاملة.

ويشكل التحليل المفاهيمي، وترجمة الوثيقة(*) التي يتم تكشيفها الجانب الفكري لعلمية التكشيف. ولكي نكون أكثر تحديدا، فإن التحليل المفاهيمي في

(*) حتى لا يثير استعمال كلمة "ترجمة" في هذا السياق شيئا من اللبس أو الغموض، نود أن نوضح أن الترجمة لا تتم للوثيقة كاملة، وإنما لعناصر المحتوى الموضوعي التي وقع عليها الاختيار، وتأتي الترجمة في مرحلة لاحقة لهذا الاختيار؛ فجميع عمليات المعالجة الموضوعية لأوعية المعلومات، سواء في التصنيف أو الفهرسة الموضوعية، أو التكشيف أو الاستخلاص، تمر بثلاث خطوات؛ أولاها الإحاطة، أي الإلمام بمختلف عناصر المحتوى الموضوعي، أي تكوين صورة شاملة كاملة لمحتوى الوثيقة. والخطوة التالية هي اختيار أو انتقاء عناصر المحتوى الموضوعي الأهم من غيرها، والجديرة فعلا بالتنويه أو الإبراز، في ناتج التحليل الموضوعي، وتسمى هذه الخطوة الترجيح. ثم تأتي خطوة الترجمة، أي تحويل الصورة التي ارتسمت في ذهن المسئول عن المعالجة الموضوعية للعناصر الجديرة بالتنويه، إلى اللغة المستعملة في التعبير عن ناتج التحليل، سواء كانت هذه اللغة خطة للتصنيف، أو قائمة لرؤوس الموضوعات، أو مكنزا. ويمكن أيضا القول بأن المعالجة الموضوعية تمر بمرحلتين؛ الأولى هي التحليل، وتشمل الإحاطة والترجيح، والثانية هي الترجمة. راجع في ذلك:

(1) Lancaster, F.W. Indexing and abstracting in theory and practice. 3rd ed. London, Library Association, 2003.

(٢) لانكستر، فردرك ولفرد. نظم استرجاع المعلومات. ط٢. ترجمة حشمت قاسم. القاهرة. مكتبة غريب، ١٩٨١.

(٣) لانكستر، فردرك ولفرد وآمي وورنر. نظم استرجاع المعلومات. ط٣. ترجمة حشمت قاسم. الرياض، مكتبة الملك فهد الوطنية، ١٩٩٧.

(٤) حشمت قاسم. مدخل لدراسة التكشيف والاستخلاص. القاهرة، دار غريب، ٢٠٠٠.

التكشيف ينطوي على التحقق من المفاهيم الرئيسة التي تحظى بالتغطية في الوثيقة وترجيح المفاهيم الجديرة بالتنويه، بينما يحوِّل الجانب الخاص بالترجمة فعلا المفهوم الذي وقع عليه الاختيار، إلى مصطلحات كشفية اعتمادا على لغة التكشيف، التي وقع عليها الاختيار سلفا. وسوف نناقش لغة التكشيف على وجه الخصوص، مع لغة الاسترجاع في الفصل الرابع.

أنواع التكشيف:

يتوقف نوع التكشيف عادة على الطريقة التي يتم بها الحصول على المصطلحات؛ فإذا كانت المصطلحات مقتطفة أو مستقاة من النص الأصلي للوثيقة، فإنه يسمى التكشيف الاشتقاقي **derivative**. وفي مقابل ذلك، إذا كانت المصطلحات معينة **assigned** للوثيقة فإنه يسمى التكشيف بالتعيين **assignment**. ومن الممكن أيضا معاملة التكشيف الاشتقاقي بوصفه مرادفا لتكشيف الكلمات المفتاحية، نظرا لأن المصطلحات الكشفية يتم التقاطها مباشرة من الكلمات المفتاحية الواردة في النص، ولا يتم الرجوع إلى أي لغة مقيدة. وفي مقابل ذلك فإنه يتعين استعمال لغة مقيدة في التكشيف بالتعيين لضمان استعمال المصطلحات الملائمة.

وعادة ما تسمى المصطلحات الكشفية التي يتم تعيينها من لغة مقيدة، بالواصفات **descriptors**، وذلك على الرغم من أن اللغة المقيدة المستخدمة قد لا تكون مكنزا. وإذا كان المفهوم الذي يتم تكشيفه جديدا، أو تصادف أن كان علما (كالسور العظيم مثلا) ولا يمكن العثور على مقابل له في اللغة المقيدة، فإنه يمكن إعداد محدد للهوية **identifier** في التكشيف بالتعيين. وبعبارة أخرى، فإن محددات

الهوية، مصطلحات يعينها المكشف، ولكنها لا تستقى من لغة مقيدة. وكما سبق أن ذكرنا، فإن المصطلحات التي يتم التقاطها في عملية التكشيف الاشتقاقي تسمى كلمات مفتاحية.

وأحيانا ما يسمى التكشيف دون الاعتماد على لغة مقيدة بالتكشيف الحر (Fugmann, 1993). ويطبق التكشيف الحر بوجه عام في التكشيف الاشتقاقي. وعلى النحو نفسه، فإن التكشيف بالتعيين عادة ما يتم باستخدام إحدى اللغات المقيدة. وكون الأمر يتطلب استخدام لغة مقيدة في التكشيف، من القضايا المثارة عبر السنين، ولا تزال المناظرة حولها مستمرة. وبوجه عام، فإنه مع تزايد توافر المعلومات رقميا، يتراجع استخدام اللغات المقيدة في التكشيف، نظرا لعوامل الكم والكيف.

التكشيف الآلي والتكشيف المعتمد على الآلات:

من الممكن تقسيم جميع الأنشطة التي ينطوي عليها التكشيف إلى فئتين؛ الأنشطة الفكرية والأنشطة الآلية **mechanical**. وقد تناولنا الجانب الفكري للتكشيف في القسم ٢ / ١ . أما الجوانب الآلية في التكشيف فتشمل في المقام الأول، أنشطة كالترتيب الهجائي وصياغة المداخل الكشفية. وبينما يظل الجانب الفكري من التكشيف مهمة بشرية في معظم الأحيان، نظرا للوضع الراهن للبحث في الذكاء الاصطناعي، فإن عمليات التكشيف ذات الطابع الآلي، يمكن أن تتم على نحو مرضٍ بالحاسبات. وإذا ما استخدمت الحاسبات في معالجة كل من الجوانب الآلية والفكرية

للتكشيف، يصبح هناك تكشيف آلي **automatic**. أما إذا ما استخدمت الحاسبات في العمليات التكرارية فقط في التكشيف، بينما يتحمل المكشفون من البشر مسئولية الجوانب الفكرية للتكشيف، يصبح هناك تكشيف معتمد على الآلات **automated**.

ويبدو التكشيف الآلي، الذي يعرف أيضا بالتكشيف بواسطة الآلات **machine indexing**، حلا له جاذبية، خاصة بالنسبة لمشكلة الاطراد والتكلفة المرتفعة المرتبطة بالتكشيف اليدوي. إلا أن الضعف الجوهري للتكشيف الآلي، يكمن في عجز الحاسب عن التعامل مع الجانب الفكري للتكشيف، إلا بمساعدة المكشف البشري، وفي معظم الأحيان يكون أداء الحاسب أسوأ بكثير من أداء اختصاصي المعلومات. والسبب في ذلك واضح؛ فالحاسب لا يستطيع في الواقع التفكير، كما أنه لا تتوافر له القدرات التحليلية التي يتمتع بها الإنسان. وفي مقابل ذلك، يمكن للتكشيف الآلي إعفاء المكشفين البشر من مهام التكشيف التكرارية المرهقة، وبذلك يكون بإمكانهم التركيز على نحو أكثر، على المهام الفكرية للتكشيف.

ولا ينبغي، من ناحية أخرى، التهوين من قدر التكشيف الآلي، وخصوصا عندما تكون كميات المعلومات المراد تكشيفها ضخمة؛ إذ عادة ما يعتمد هذا الضرب من التكشيف على خوارزميات متعددة، من بينها تبدو طريقة إحصاء تردد الكلمات المفتاحية الأكثر شيوعا، وقد استخدمت فعلا بكثافة. وتستند طرق التكشيف الآلي الأخرى بوجه عام، على تقارب الكلمات المفتاحية، وموقعها، والاحتمالات، والخصائص اللغوية. كما تدخل اللغات المقيدة في بعض تطبيقات التكشيف الآلي أيضا، إلا أن نجاح مثل هذه الطريقة يتأثر سلبا بالطابع الآلي للتكشيف الإلكتروني.

التكشيف في بيئة التنظيم الفائق:

يتزايد توافر المعلومات في بيئة التنظيم الفائق **hyperstructure**، التي تتمثل في العنكبوتية العالمية **World Wide Web**. وبالنسبة للمعلومات المعتمدة على العنكبوتية، فإنه من الطبيعي إلى أبعد حد، تقديم المصطلحات الكشفية بوصفها روابط فائقة **hyperlinks**، تتضمن كلا من المصطلحات الكشفية وتقنية تحديد المكان **locator mechanism**. وبعبارة أخرى، فإن أسماء الروابط الفائقة تقوم مقام المصطلحات الكشفية، حيث تقود طريقة الروابط الفائقة المستفيد بسلاسة، إلى حيث تشير المصطلحات الكشفية. وبالمقارنة ببيئات التكشيف الأخرى، فإن هذا النظام فريد من النواحي التالية:

أولا، ترد المصطلحات الكشفية في بيئة التنظيم الفائق كامنة في ثنايا الوثائق نفسها، وليست ككيانات منفصلة خارج السياق. ثانيا، تندمج المصطلحات الكشفية وتقنية الوصول إلى المكان معا في وحدة واحدة، بدلا من ورودهما منفصلين. ثالثا، لا ينعكس التقسيم الهرمي الموضوعي، أو مفهوم المداخل الفرعية في التكشيف التقليدي، تماما أو بحذافيره في بيئة التنظيم الفائق. رابعا، لا يمكن النظر إلا للروابط القائمة على المحتوى، في التنظيم الفائق، بوصفها مصطلحات كشفية، بينما لا يمكن للروابط التنظيمية، "كالصفحة التالية" و"رأس الوثيقة" أن تعد كذلك **(Chu, 1997)**. خامسا، ينهض المؤلفون بدور المكشف عندما يقومون بإعداد وثائق التنظيم الفائق، كما يتم التكشيف تزامنيا مع إنشاء الوثائق، إن لم يكن قبل ذلك. ويترتب على

ذلك احتمال كثافة ورود مصطلحات (من قبيل "اضغط هنا" على سبيل المثال) تلك المصطلحات التي لا يحتمل لها أن يقع عليها الاختيار من قبل المكشف، بوصفها أسماء روابط، في بيئة التنظيم الفائق. سادسا، تراجع احتمالات التفاوت بين الوثائق الأصلية والمصطلحات الكشفية، في طريقة تنظيم المعلومات هذه؛ نظرا لأن المؤلف هو من يقرر، أثناء كتابة الوثيقة، أي المصطلحات ينبغي أن تصبح روابط، بينما يحدث في التكشيف التقليدي أن يكتب المؤلف الوثيقة، ثم يحاول المكشف بعد ذلك تحليلها لأغراض التنظيم.

ونظرا للخصائص الفريدة للتكشيف في بيئة التنظيم الفائق، فإن الوثائق الفائقة **hyperdocuments** ينبغي أن تكشف بطرق تلائمها. فعلى المرء، على سبيل المثال، أن يختار أسماء الروابط بعناية، عند كتابة الوثائق الفائقة **(Chu, 1997)**.

التمييز الاجتماعي:

بدأ التمييز الاجتماعي **social tagging**، الذي يختلف تمام الاختلاف عن التمييز بإحدى شفرات التحقق من الهوية **identification code** (كالتحقق بالتردد اللاسلكي **radio frequency identification - RFID** على سبيل المثال) يزداد انتشارا وشيوعا، منذ عام ٢٠٠٣، بوصفه وسيلة بالنسبة للمستفيد النهائي، من أجل "تكشيف" المعلومات على الإنترنت. وينتمي التمييز الاجتماعي، مرتبطا بالتطبيقات الأخرى الرامية إلى مشاركة المستفيد (كالتدوين **blogging**، والمواقع الجماعية على العنكبوتية **Wikis**) إلى حركة الجيل الثاني للعنكبوتية **Web 2.0** **(O'Reilly,**

(2005. وتشمل النظم المبكرة المزودة بتقنية تمييز فلكر **Flickr**، وهو موقع شهير على العنكبوتية لتقاسم الصور الضوئية، و**del.icio.us** الذي تغير اسم نطاقه إلى **Delicious.com** عام ٢٠٠٧. ففي مثل مواقع التمييز هذه، يمكن للمرء أن يصف أو يعقب على ما يراه أو يسمعه، بتمييزه بكلمات أو عبارات من اختياره.

والتمييز، إلى حد بعيد، نشاط مماثل لتكشيف الكلمات المفتاحية الذي سبقت الإشارة إليه في هذا القسم، إلا أن تكشيف الكلمات المفتاحية يتم إجراؤه بواسطة المكشفين، ومؤلفي الوثائق التخصصية، ونظم استرجاع المعلومات، بينما يتم إجراء التمييز بواسطة أي مستفيد من الإنترنت. ونظرا لهذا الاختلاف الجوهري، فإن التمييز لا يبدو مقابلا لتكشيف الكلمات المفتاحية، وإن كان من يمارسونه **taggers** يميلون لاختيار أسماء أو عبارات اسمية كتيجان **tags**، في معظم الأحيان **(Kipp, 2007; Xu & Chu, 2008)**. يضاف إلى ذلك وجود المشكلات التي يعاني منها تكشيف الكلمات المفتاحية (كالمترادفات، والمشترك اللفظي أو الجناس، وتعدد أشكال الكلمات، على سبيل المثال) جميعها في التمييز، إن لم تكن على نطاق أوسع. إلا أن للتمييز، كطريقة ناشئة وبديلة للتعبير عن المعلومات أو تمثيلها، احتمالات عريضة في تيسير استرجاع المعلومات، نظرا لأن التيجان التي ينشئها المستفيدون، من بين مزاياها الأخرى (كتوفير المزيد من نقاط الوصول على سبيل المثال)، من المرجح إلى حد بعيد، أن يتم اختيارها من قبل المستفيدين النهائيين أنفسهم كمصطلحات استفسار للعثور على المعلومات.

وكأسلوب ناشئ لتنظيم المعلومات أو تمثيلها، حقق التمييز ابتكارات من شأنها إثـراء مجـال اسـترجاع المعلومـات (Smith, 2008). والتـصنيف المتعـارف عليـه **folksonomy** من بين هذه الابتكارات. وقـد سـك تومـاس فانـدر وال **Thomas Vander Wal** كلمة **folksonomy** عام ٢٠٠٤، وهي تركيب مزجـي مـن كلمتـي **folks** و**taxonomy** (Vander Wall, 2007). وبعبـارة أخـرى فـإن التـصنيف المتعارف عليـه **folksonomy** تـصنيف شـجري أو هرمـي **taxonomy**، قـائم عـلى التيجان التي ينشئها القومَ **folks** أو المستفيدون النهائيون. ويشتمل القسم ٤/ ٤ عـلى مزيد من المناقشة لكل من التصنيفات الشجرية والتصنيفات المتعارف عليها. وعادة ما تتخذ التصنيفات المتعارف عليها شكل حشد من التيجان **tag cloud**، وهو تجمع من التيجان التي تنشأ في موقع تمييز ما، اعـتماداً عـلى الـترددات المميـزة. ويمكـن لحـشد التيجان أن يساعد المستفيدين في اختيار المصطلحات الملائمة سواء في عملية التمييز أو في عملية الاسترجاع.

التقسيم الفئوي:

يـدل التقـسيم الفئـوي **categorization** عـلى التنظـيم التسلـسلي والهرمـي للمعلومات أو تمثيلها بفئات. ويمكن أن يكون هناك ضربان من التقسيم الفئوي، تبعا لما إذا كانت هناك خطة نسقية متطورة مستخدمة أم لا.

أنواع التقسيم الفئوي:

عادة ما يتم التقسيم الفئوي وفقـا لخطـة تـصنيف مـستقرة، كتـصنيف ديـوي العشري (DDC) وتصنيف مكتبة الكونجرس (LCC). وتسمى تطبيقـات التقـسيم

الفئوي هذه على هذا النحو بوجه عام، بالتصنيف الذي يمكن أن يطبق على وجه العموم على مجموعات المكتبات ومرافق المعلومات. وتشمل رموز التصنيف الأرقام، أو الهجائيات، أو الأرقام الهجائية.

ومع ظهور الإنترنت، نشر الكثير من المعلومات على هذه المنصة الجديدة. وتتقاسم المعلومات على الإنترنت بوجه عام الخصائص التالية، وهي: الطابع المؤقت أو سرعة الزوال **ephemeral**، والجودة المتفاوتة **mixed quality**، والكميات الهائلة. ولهذا فإن طريقة التصنيف تبدو باهظة التكلفة، أو غير ملائمة، أو غير مجدية لتنظيم المعلومات أو تمثيلها على الإنترنت، بينما يبدو التقسيم الهرمي أو الشجري القائم على فئات غير محكمة البنيان، الإطار المناسب لتنظيم المعلومات من هذا الطابع. وياهو! **Yahoo!** ولاشك، رائد في هذا المجال، إذ أصبح أنموذجا لتنظيم المعلومات أو تمثيلها على الإنترنت، وعلى العنكبوتية العالمية على وجه الخصوص. وترد الفئات في التصنيفات الشجرية من أجل تنظيم المعلومات المعتمدة على العنكبوتية، أو تسجل مباشرة بوصفها روابط فائقة. ولا تستخدم أي رموز **notations** (كالأرقام أو الهجائيات أو الأرقام الهجائية مثلا) للدلالة على العلاقات الشجرية أو غيرها في إطار التصنيف. ونقدم المزيد من التغطية للتصنيف في الفصل الرابع، إلى جانب مناقشة التصنيفات المتعارف عليها **folksonomies** والتصنيفات التخصصية **ontologies**.

أسس التقسيم الفئوي:

عندما يقع الاختيار على التقسيم الفئوي، كطريقة لتنظيم المعلومات أو تمثيلها، يتم تخصيص فئة واحدة أو فئتين أحيانا للعناصر التي تغطي موضوعات بينية أو متعددة الارتباطات **interdisciplinary** في وثيقة معينة. وبعبارة أخرى، فإن وثيقة

بعينها ينبغي أن تعين لفئة محددة واردة في النظام القائم للتقسيم الفئوي. ويتطلب ذلك أن تكون الفئات التي وقع عليها الاختيار لنظام التقسيم الفئوي في الوقت نفسه:

- شاملة **Exhaustive**.
- قائمة بذاتها **Mutually exclusive**.

ويقصد هنا القول بأن جميع الفئات التي يمكن أن تدعو الحاجة إليها لتنظيم المعلومات أو تمثيلها، ترد في نظام التقسيم الفئوي. ولا ينبغي في الوقت نفسه لأي من هذه الفئات أن تكون متداخلة فيما بينها. وما لم يتوافر الشرط الأول فإنه لن يكون من الممكن التعبير عن بعض المعلومات أو تنظيمها بالفئات المتاحة في الخطة. وما لم يراعى الشرط الثاني فإنه يمكن حينئذ اختيار أكثر من فئة واحدة لتمثيل الوثيقة الواحدة. ومن شأن أسس أخرى كالمرونة مثلا، أن تيسر الاستخدام. وعلى الرغم من أهمية المرونة، فإنها ليست جوهرية كهذين الأساسين السابقين.

اندماج طريقتي التقسيم الفئوي:

يكمن ما يميز أياً من طرق التقسيم الفئوي عن غيرها، في المقام الأول، في طبيعة الإطار المتبع لأغراض تنظيم المعلومات أو تمثيلها. فقد أكد التصنيف، وقد صمد لأنواع متعددة من الاختبارات، مكانته كطريقة بارزة لتنظيم المعلومات. أما التقسيم الشجري، فعلى العكس؛ إذ ينظر إليه بوصفه طريقة سريعة غير ناضجة لتنظيم المعلومات. ومع تزايد توافر المعلومات المهمة الجيدة على الإنترنت، بدأ التصنيف يستخدم للتقسيم الفئوي للمعلومات المعتمدة على الشبكة

(McKierman, 2001 على سبيل المثال). وفي الوقت نفسه تم تنقيح الأطر، وتطويرها لأغراض التقسيم الشجري والارتقاء بها، بإدخال الخصائص التي صمدت لاختبار الزمن، تلك الخصائص التي نشأت في كنف خطط التصنيف، كالتنظيم أو التقسيم الشجري.

وفضلا عن ذلك، تزايدت جهود استكشاف التقسيم الفئوي للنصوص بالوسائل الآلية (Boley, et al., 1998; Yang, 1999 على سبيل المثال) وذلك بالاتجاه نحو التقسيم الشجري **taxonomy** لا التصنيف **classification**(*)، مع تزايد كم المعلومات التي تتاح رقميا. وبينما اجتذب التصنيف الآلي يوما ما، قدرا كبيرا من اهتمام الباحثين، فإن القدرات العقلية اللازمة للتصنيف المنضبط، تبدو صعبة التحقيق في الخوارزميات القائمة على الآلات دون سواها.

التلخيص:

على عكس التكشيف أو التصنيف أو التقسيم الفئوي، فإن التلخيص يحاول التعبير عن وثيقة ما أو تمثيلها بفقرة أو اثنين، أي بملخص أو شكل مركز للأصل. وتحدد درجة التركيز والطريقة التي تتبع معالم الطرق المختلفة للتلخيص.

(*) أيا كان المصطلح المستعمل للدلالة على التقسيم الفئوي، فإن التقسيم الفئوي هو الأساس في المعالجة الموضوعية للوثائق أو أوعية المعلومات؛ فالترتيب الهجائي للمفردات تقسيم فئوي، والفهرسة الموضوعية تقسيم فئوي لأنها تؤدي إلى نشأة فئات من الوثائق أو أوعية المعلومات، وكذلك التكشيف، أيا كان شكله وتقنياته، تقسيم فئوي لأنه يؤدي أيضا إلى نشأة فئات، بما في ذلك تكشيف الاستشهاد المرجعي **citation indexing**. (المترجم)

فئات التلخيص:

المستخلصات:

المستخلص **abstract** تمثيل موجز دقيق لمحتوى الوثيقة. ولقد كان الاستخلاص، أي عملية صياغة المستخلصات، يتم أساسا بواسطة البشر، على الرغم من بذل بعض الجهود الآلية (**Luhn, 1958** على سبيل المثال) فيما مضى. فالمستخلصات ينبغي مثاليا أن تصاغ بأسلوب مماثل للأصل، إلا أن ذلك المعيار من الصعب الوفاء به، نظراً للفقدان الذي لا مناص منه للأصالة أو الموثوقية، أثناء عملية الاستخلاص.

ومن الممكن أيضا تقسيم المستخلصات إلى ثلاثة أنواع؛ المستخلصات الإعلامية أو المعلوماتية **informative**، والمستخلصات الكشفية **indicative**، والمستخلصات النقدية **critical**. وتشتمل المستخلصات المعلوماتية على المعلومات الجوهرية الواردة بالوثيقة الأصلية، ولهذا فإنها يمكن أن تقوم مقام البديل. واعتمادا على المستخلصات المعلوماتية يمكن للمرء أن يقرر ما إذا كان بحاجة لأن يطلع على الوثيقة الكاملة. أما المستخلصات الكشفية فإنها يمكن في مقابل ذلك أن تصف "حيثيات أو محتوى **aboutness**" الوثيقة الأصلية، مع استبعاد تفصيلات معينة كالمنهج والنتائج، ومن ثم فإنها لا يمكن أن تعامل بوصفها بديلا عن الأصل. وعلى المرء الاطلاع على الوثيقة الأصلية، للحصول على تصور شامل لها. وتسمى المستخلصات النقدية على هذا النحو؛ نظرا لأنها لا تمثل محتوى الوثيقة موضوع الاهتمام فحسب، وإنما تحاول أيضا تقييم ذلك المحتوى. ولا يوصى بهذا الشكل من المستخلصات؛ نظرا لأن المستخلصات بطبيعتها، ينبغي أن تخلو من أي تأويل أو تفسير أو أحكام. ولا ينبغي بوجه عام كتابة مستخلصات نقدية ما لم يطلب من المرء ذلك.

وكما سبق أن أشرنا، فقد استكشف الباحثون طرق إعداد المستخلصات الآلية، إلا أن الناتج النهائي يبدو الآن في جوهره، أقرب للملخصات الآلية أو المقتطفات الآلية، منه إلى المستخلصات الآلية، لأنه يتكون في المقام الأول، من جمل مفتاحية مقتطفة أو مقتبسة من الوثائق الأصلية.

الملخصات:

الملخص إعادة صياغة للنقاط الرئيسة الواردة في الوثيقة الأصلية، يرد في صدر الوثيقة أو في نهايتها. وعلى الرغم من قربه من المستخلص، فإن الملخص عادة ما يفترض أن القارئ سوف تتاح له الفرصة لإمعان النظر في الوثيقة الكاملة. ومن ثم، فإن هناك عناصر معينة، لا غنى عنها للاستيعاب الكامل للوثيقة، كالأقسام المتعلقة بالخلفيات أو المنطلقات، والهدف، والمنهج، عادة ما تغيب من الملخص **(Rowley, 1992)**.

ولقد تطورت خوارزميات لتلخيص النصوص آليا، وخصوصا تلك النصوص التي تتخذ الشكل الرقمي، في غضون السنوات القليلة الماضية (**Spärck Jones, 2007; Spärck Jones & Endres-Niggemeyer, 1995** على سبيل المثال). ويبرز التلخيص الآلي، في الواقع، بوصفه مجالا للبحث الجاري في الذكاء الاصطناعي **(AI)**. ويمكن لبعض الباحثين تسمية منتجاتهم النهائية، من البدائل أو التمثيلات بالمستخلصات. إلا أنه من المعروف ولاشك، أنه لا يمكن إلا للبحث في الذكاء الاصطناعي، أن يكون قادرا على أن يجعل من الاستخلاص الآلي حقيقة واقعة، ولم يتم تطوير خوارزميات الذكاء الاصطناعي هذه بنجاح. ونواصل مناقشة التخليص الآلي في القسم ١٢/ ٢/ ٢.

المقتطفات:

يتكون المقتطف **extract** من جزء واحد أو أكثر من وثيقة ما، يتم التقاطه لتمثيل الكل. ولا يمكن للمقتطفات أن تكفل تمثيلا مناسبا للوثائق الأصلية، إلا أنها يمكن أن تكون لها أهميتها بالنسبة للقارئ الذي يهتم بحصيلة أو خلاصة دراسة معينة. ولا يمكن بحال النظر إلى المقتطفات بوصفها بدائل للوثائق الأصلية.

والمقتطفات، وإن كانت قدرتها التمثيلية أدنى من تلك الخاصة بكل من الملخصات والمستخلصات، من الممكن بناؤها كاملة بالطرق الآلية. وتستخدم معظم نظم استرجاع الإنترنت، بما في ذلك جوجل، مقتطفات آلية لأغراض تمثيل المعلومات. ومن الخصائص السيئة للمقتطفات الآلية التي تنتجها نظم استرجاع الإنترنت، أن المقتطف عادة ما ينتهي بعلامات الحذف **ellipsis** عندما يصل إلى نقطة التوقف **cut off point** التي يحددها النظام. من ثم فإن جودة المقتطفات الآلية أمر يثير المخاوف أو لا يدعو للاطمئنان، عندما تستخدم لتمثيل الوثائق.

ملخص الموقع الثري:

آر إس إس **RSS**، الشكل المختصر الذي يقوم مقام ملخص الموقع الثري **Rich Site Summary**، تطبيق آخر للجيل الثاني للعنكبوتية العالمية **Web 2.0**، يمكن أن يستخدم لأغراض تمثيل المعلومات. وبمزيد من التحديد، فإن ملخص الموقع الثري يعتمد على صيغ ملف يسمى ملقمات **feeds**، يقوم بتجميع المعلومات التي استجدت حديثا من عدة مصادر على العنكبوتية. وبإمكان أولئك الذين يشتركون في ملقمات معينة، عن طريق قارئ ما (أي برنامج تجميع **aggregator**) أن

يتلقوا، على حاسباتهم، المعلومات التي تتيحها حديثاً هذه الملقيات. ومن ثم فإنه يمكن النظر إلى ملخص الموقع الثري بوصفه خدمة إحاطة جارية، في بيئة العنكبوتية العالمية، تقدم للمشتركين ملخصات للمعلومات التي تتاح حديثا، من المصادر التي يهتمون بها. وبهذا المعنى يبدو ملخص الموقع الثري اسما ملائما لهذا التطبيق.

وإدراكا وتقديرا منهم لأهمية ملخص الموقع الثري، ابتكر أولئك الذين ينهضون بمهام في تجمع العنكبوتية العالمية **World Wide Web (W3C) Consortium**، طبعة جديدة من ملخص الموقع الثري، عندما أصبحت نتسكيب **Netscape**، الشركة التي قامت بتطوير أول متصفح تصويري للعنكبوتية **graphic web browser**، عاجزة عن دعم ملخص الموقع الثري الذي أنشأته من البداية. ونظرا لأن الطبعة الجديدة من ملخص الموقع الثري كانت تستند إلى المواصفة المعيارية المسماة بإطار وصف المصادر **Resource Description Framework (RDF)**، التي وضعها تجمع العنكبوتية العالمية **(W3C)**، كجزء من الهندسة المعمارية للعنكبوتية الدلالية؛ تغير المختصر **RSS** للدلالة على ملخص موقع آر دي إف **RDF Site Summary**، لتمييزه عن سلفه ملخص الموقع الثري **(Kelly, 2005)**. والتوظيف أو التطبيق البسيط فعلا **Really Simple Syndication** معنى آخر للشكل المختصر **RSS**، تم تطويره بعد طبعة تجمع العنكبوتية العالمية **W3C** **(Lerner, 2004)**.

ومقارنة بطرق التخليص الأخرى التي تناولناها في هذا القسم، فإن ملخص الموقع الثري يتم إنجازه آليا على العنكبوتية. وتلبي هذه الخاصية الآلية على نحو ملائم، احتياجات كثير من المستفيدين النهائيين، الذين يرغبون في الاطلاع على أحدث المعلومات، من ناحية، إلا أن جودة معلومات الملخص الذي يقدمه ملخص الموقع

الثري، من ناحية أخرى أدنى مستوى، ولاشك من معلومات طرق التخليص الأخرى، كالمستخلصات على سبيل المثال. والقدرة التمثيلية **representativeness** أحد المعايير التي عادة ما تستخدم في الحكم على جودة التلخيص.

قضية القدرة التمثيلية:

جميع الأشكال الأربعة المختلفة للتلخيص، المستخلصات، والملخصات، والمقتطفات، وملخص الموقع الثري، وسائل معترف بها لتمثيل المعلومات. إلا أنها كما سبق أن بينا، تختلف فيما بينها من حيث مدى التمثيل **representativeness**. وبترتيب هذه الوسائل الأربعة بناء على هذا المعيار، يمكن للمستخلصات أن تتربع على القمة، بينما يبدو ملخص الموقع الثري متمتعا بأقل قدرة تمثيلية في هذه المجموعة. ويمكن للملخصات والمقتطفات أن ترد في مرتبة وسط بين الوسيلتين الأخريين. وعلى الرغم من إيجاز هذه المقارنة، فإنها تلقي بعض الضوء على سبل تطبيق أي من طرق التلخيص، على أفضل وجه، لمختلف مهام تمثيل المعلومات.

الطرق الأخرى لتنظيم المعلومات:

إذا ما نظرنا إلى التكشيف والتقسيم الفئوي والتلخيص، بوصفها طرقا تقليدية لتمثيل المعلومات أو تنظيمها، فإن الأساليب التي تحظى بالمناقشة في هذا القسم فريدة وغير معيارية.

الاستشهادات المرجعية:

الاستشهادات المرجعية **citations** هي الإشارات أو الإحالات **references** التي يسجلها المؤلفون عندما يؤلفون أعمالهم. ويدل الاستشهاد المرجعي بوجه عام، على وجود علاقة بين جزء من الوثيقة المستشهد بها أو كلها، وجزء من الوثيقة التي يرد بها الاستشهاد أو كلها **(Malin, 1968)**. ومنذ أن أدخل يوجين جارفيلد **Eugene Garfield**، مؤسس معهد المعلومات العلمية **Institute for Scientific Information**، الذي ينشر كشافات الاستشهاد المرجعي **citation indexes**، هذا النهج الجديد إلى مجال تنظيم المعلومات واسترجاعها، تركز قدر كبير من المناقشات على فك شفرة دلالة الاستشهاد المرجعي، أو التحقق من أسباب تسجيل الاستشهادات. وحاول جارفيلد عام ١٩٦٥، تلخيص أسباب تسجيل الاستشهادات المرجعية، في خمس عشرة نقطة، تشمل على سبيل المثال الإعراب عن الولاء للرواد، والاعتراف بالأعمال ذات الصلة بالعمل الجديد **(Garfield, 1965)**(*).

وعلى الرغم من أن الدافع الحقيقي للاستشهاد بأعمال الآخرين، قد يختلف من شخص لآخر، فإنه من الممكن النظر إلى الاستشهادات المرجعية، بوصفها ما يختاره المؤلفون الذين يسجلون الاستشهادات المرجعية، لتمثيله في أعمالهم. ويتخذ التمثيل

(*) راجع أيضا في ذلك:

- حشمت قاسم. كشافات الاستشهاد المرجعي وإمكاناتها الاسترجاعية. المجلة العربية للمعلومات، مج٢، ع٤؛ يونيو ١٩٨٠. وأعيد النشر في: حشمت قاسم. دراسات في علم المعلومات. ط٢. القاهرة، دار غريب، ١٩٩٥، ص ص ٣٣٥-٣٥٣.
- حشمت قاسم. تحليل الاستشهادات المرجعية وتطور القياسات الوراقية. المجلة العربية للمعلومات، مج٣، ع٥؛ ديسمبر ١٩٨٠. وأعيد النشر في: حشمت قاسم. دراسات في علم المعلومات. ط٢. القاهرة، دار غريب، ١٩٩٥، ص ص ١١٧-١٤٧.

في هذا السياق شكل الاستشهادات المرجعية، لا البدائل العادية كالمستخلصات والمصطلحات الكشفية. والاستشهادات المرجعية، في الحقيقة معلومات وراقية (ببليوجرافية) حول الوثائق، يمكن تحليلها إلى عناصر كالعناوين وأسماء المؤلفين. وبعبارة أخرى، فإن الاستشهادات المرجعية تكفل مصدراً ثرياً لتنظيم المعلومات. وفضلا عن ذلك لا تدعو الحاجة إلى ابتكار إطار (كالمكنز أو خطة التصنيف على سبيل المثال) لأغراض التنظيم، بينما لا يمكن الاستغناء عن مثل هذا الإطار في التصنيف والتكشيف باللغات المقيدة(*).

والمؤلفون الذين يسجلون الاستشهادات المرجعية، إنما هم في الواقع يكشفون أعمالهم بمصادر متاحة بسهولة، وهي على وجه التحديد الإشارات أو الاستشهادات المرجعية. إلا أن الاستغناء عن الوسطاء في تكشيف الاستشهادات المرجعية، له آثاره الإيجابية والسلبية على تنظيم المعلومات أو تمثيلها؛ أما التأثير الإيجابي فهو أن المؤلف بوصفه المكشف، يعرف عمله على أحسن وجه، ولا يحتاج إلى بذل أي جهد لتفسير الوثائق الأصلية، وأما التأثير السلبي فهو أن المبرر الحقيقي للاستشهاد المرجعي غير معروف في الواقع. ولهذا فإننا يمكن أن نتساءل دائما ما إذا كانت الاستشهادات المرجعية المناسبة قد استخدمت فعلاً لتمثيل وثيقة ما. وهناك داع آخر لأن نتخوف من تكشيف الاستشهادات المرجعية بوصفه طريقة لتنظيم المعلومات، وهو قصور تغطية مراصد بيانات الاستشهاد المرجعي. فالباحثون عادة ما يفيدون من مثل هذه المراصد، لأن إعداد مرصد من الصفر غالبا ما يبدو أمرا غير عملي ومبدداً للوقت إلى حد بعيد.

(*) ربما كان من الممكن هنا القول بأن هناك فئتين من طرق تنظيم المعلومات؛ الطرق الصريحة **Explicit** المألوفة كالتصنيف والفهرسة الموضوعية والتكشيف، والطرق الضمنية **implicit** كالاستشهاد المرجعي. (المترجم)

ونظرا لأن تكشيف الاستشهادات المرجعية يتطلب القليل من الذكاء البشري، فإنه يمكن للعملية برمتها أن تتم آليا بسهولة. والواقع أنه يمكن لعملية تكشيف الاستشهادات المرجعية أن تتم كاملة دون أي تدخل بشري، الأمر الذي يبدو غير قابل للتحقيق بالنسبة للطرق الأخرى لتنظيم المعلومات.

التسلسلات:

بينما يمكن للاستشهادات المرجعية أن تقوم مقام البدائل بالنسبة للوثائق الأصلية، فإن التكشيف التسلسلي **String indexing (Craven, 1986)**(*) نهج آخر لتنظيم المعلومات. ويمكن للتسلسلات أن تكون عبارات، أو مجموعات من العبارات، أو تصريحاً **statement**، ويتم إعدادها جميعا لتمثيل وثيقة ما على وجه التحديد "فاستخدام العنكبوتية العالمية للتعلم اللامركزي في التعليم العالي **Use of world wide web for distributed learning in higher education**"، مثال لهذه التسلسلات.

وهناك أنواع مختلفة كثيرة من الكشافات التسلسلية، إلا أنها تتقاسم فيما بينها خاصتين رئيستين: (١) صياغة سلسلة لتلخيص موضوع وثيقة ما يدويا، و(٢) توليد مداخل كشفية آليا اعتمادا على التسلسل المحدد لأغراض التنظيم. ولهذا فإن التكشيف التسلسلي حالة متميزة للتكشيف الآلي، الذي عولج بإيجاز في ٢/ ١/ ٢.

(*) استعملت المؤلفة هنا كلمة **strings**، مرادفا لكلمة **chains** التي ارتبطت تقليديا بهذا الشكل من الكشافات **chain indexes**. (المترجم)

ونظام كشاف السياق المصون **PRE served context index system** (بريسيز **PRECIS**)، ونظام تكشيف العبارات المسورة **Nested Phrase Indexing System** (نفيز **NEPHIS**)، اثنان من أشهر النظم التي تطورت للتكشيف التسلسلي؛ ففي كلا النظامين، يتم أولا إعداد تسلسلات بواسطة مكشف بشري. وكشاف البريسيز مستخلص مصغر محمَّل بمصطلحات التكشيف المناسبة، وكشاف نفيز عبارة موجزة لتكشيف وثيقة ما. ثم يتم بعد ذلك تشفير أو ترميز **coded** هذه التسلسلات لبيان أي المصطلحات يمكن أن تكون نقاط وصول، وإقرار مكونات المداخل الكشفية. ويمكن لخوارزميات العرض الفعلية، التي تستخدم في كل نظام من نظم التكشيف التسلسلي، أن تختلف اختلافا بينا. إلا أن جميع المهام الآلية في التكشيف التسلسلي، يمكن معالجتها على نحو فعال بالوسائل الآلية.

ويؤدي تكامل التحديد البشري لتسلسلات التكشيف، مع المعالجة التي تتم بواسطة الحاسب لأشكال عرض الكشافات، إلى جعل التكشيف التسلسلي طريقة لتنظيم المعلومات لها جاذبيتها. فمثل هذا التكامل يحافظ على الجودة المرتبطة بالتكشيف البشري، إذ يتكفل البشر بإعداد تسلسلات التكشيف، من ناحية، ومن ناحية أخرى يمكن للتكشيف التسلسلي أن يحول دون الاستياء المرتبط بالطرق اليدوية المثيرة للضجر غير المطردة المفتقرة إلى الكفاءة. ومهما يكن، فإن العصر الرقمي قادم بسرعة، ومن المتوقع للحاسبات أن تنهض بدور متزايد الأهمية، في تنظيم المعلومات واسترجاعها على السواء.

مراجعة للطرق الرئيسة لتنظيم المعلومات:

هناك، كما تبين لنا في الأقسام السابقة، طرق متعددة لتنظيم المعلومات. ويلخص الجدول رقم ٢/ ١ طرق التنظيم الرئيسة التي ناقشناها حتى الآن، من حيث نوع التنظيم،

والكيانات التي يتم تنظيمها، وإطار التنظيم (ما إذا كان يعتمد على لغة مقيدة) فضلا عن طريقة الإنتاج. ولكل من هذه الطرق مظاهر قوتها ومواطن ضعفها. وعندما يقع الاختيار على طريقة بعينها لتنظيم المعلومات، كما يبين ميدو **Meadow (1992)**، فإنها ينبغي أن تحقق ما يلي:

(١) تميز بين مختلف الكيانات، (٢) تتحقق من الكيانات المتشابهة، (٣) تكفل الوصف الدقيق للكيانات، (٤) تحد من الغموض في التفسير. **(pp. 41-43)**

الجدول رقم ٢/ ١ طرق تنظيم المعلومات

الطريقة / الخاصية	التكشيف		التقسيم الفئوي		التلخيص		أخرى	
نوع التنظيم	اشتقاقي	تعين	تصنيف	تقسيم شجري	مستخلص ملخص	مقتطف	استشهاد مرجعي	سلسلة
الكيان المنظم	جزء		كل		كل	جزء	كل	
إطار التنظيم	لا	نعم	نعم	لا	لا		لا	ربما
طريقة الإنتاج	آليا	يدويا وآليا	يدويا	يدويا وآليا	يدويا	يدويا وآليا	آليا	بمساعدة الآلات

وكما هو متوقع، فإنه لا يمكن لطريقة بعينها أن تلبي الشروط التي حددها ميدو كاملة. ويمكن لموطن الضعف في طريقة ما أن يكون مظهر قوة في غيرها؛ فالتكشيف وحده على سبيل المثال، لا يمكن أن يكشف عن جوانب معينة للوثيقة، بينما يقصد بالاستخلاص تمثيل وثيقة ما بكاملها. ولهذا، فإن الجمع بين الطرق (أي استخدام أكثر من طريقة واحدة) هو الخيار، عندما يحتاج المرء للاختيار من بين الطرق المتنوعة لتنظيم المعلومات. وفي التطبيق العملي، يتم استخدام مجموعة من طرق التنظيم (كالتصنيف مثلا مع الاستخلاص، أو التلخيص مع التكشيف). ويتزايد استخدام الطرق الناشئة كملخص الموقع الثري **RSS** مثلا، والتمييز **tagging**، لتلبية الاحتياجات المتغيرة لتنظيم المعلومات في العصر الرقمي.

المراجع

Boley, Daniel, et al. (1998). A client-side web agent for document categorization. Internet Research: Electronic Networking Applications and Policy, 8(5), 387-399.

Chu, Heting. (1997). Hyperlinks: How well do they represent the intellectual content of digital collections? Proceedings of the 60th Annual Meeting of the American Society of Information Science, 34, 361-369.

Craven, Timothy C. (1986). String indexing. New York: Academic Press.

Fugmann, Robert. (1993). Subject analysis and indexing: Theoretical foundation and practical advice. Frankfurt/ Main, Germany: Indeks Verlag.

Garfield, Eugene. (1965). Can citation indexing be automated? In Mary E. Stevens, et al. (Eds.), Statistical methods for mechanized documentation (pp. 189-192). Washington, DC: National Bureau of Standards.

Kelly, Brian. (2005). RSS: More than just news feeds. New Review of Information Network, 11(2), 219-227.

Kipp, Margaret E.I. (2007). @toread and cool: Tagging for time, task and emotion. Proceedings of the 8th Information Architecture Summit. Las Vegas, Nevada, March 22-27, 2007. Retrieved January 19, 2009, from eprints.rclis.org/10445/1/mkipp-iasummit2007.pdf.

Lerner, Reuven M. (2004). At the forge: Syndication with RSS. Linux Journal, 2004(126), 8. Retrieved November 12, 2008, from www.linuxjournal.com/article/7702.

Lesk, Michael. (1997). Practical digital libraries: Books, bytes and bucks. San Francisco: Morgan Kaufmann.

Luhn, H. P. (1958). The automatic creation of literature abstracts. IBM Journal of Research and Development, 2(2), 159-165.

Malin, Morton V. (1968). The Science Citation Index: A new concept in indexing. Library Trends, 16, 374-397.

McKiernan, Gerry. (2001). Beyond bookmarks: Schemes for organizing the web. Retrieved January 19, 2009, from www.public.iastate.edu/-CYBERSTACKS/CTW.htm

Meadow, Charles T. (1992). Text information retrieval systems. San Diego: Academic Press.

O'Reilly, Tim. (2005). What is Web 2.0? Retrieved October 4, 2009, from oreilly.com/web2/archive/what-is-web-20.html

Rowley, Jennifer E. (1992). Organizing knowledge: An introduction to information retrieval. Brookfield, VT: Gower.

Smith, Gene. (2008). Tagging: Emerging trends. Bulletin of the American Society for Information Science and Technology, 34(6), 14-17.

Spärck Jones, Karen. (2007). Automatic summarizing: the state of the art. Information Processing & Management, 43(6), 1449-1481.

Spärck Jones, Karen, and Endres-Niggemeyer, Brigitte (Eds.). (1995). Summarizing text [Special issue]. Information Processing & Management, 31(5).

Vander Wal, Thomas. (2007). Folksonomy. Retrieved November 20, 2008, from vanderwal.net/folksonomy.html

Xu, Chen, and Chu, Heting. (2008). Social tagging in China and the USA: A comparative study. Proceedings of the Annual Meeting of the American Society for Information Science and Technology. Columbus, Ohio, October 24-29, 2008 [CD-ROM].

Yang, Yiming. (1999). An evaluation of statistical approaches to text categorization. Information Retrieval, 1(1/2), 69-90.

الفصل الثالث

تنظيم المعلومات ٢: الموضوعات ذات الصلة

ناقش الفصل الثاني الطرق الرئيسة لتنظيم المعلومات أو تمثيلها، مع التركيز على التطبيقات التقليدية وتنويعاتها وامتداداتها. ونستكشف في هذا الفصل الموضوعات الأخرى المتصلة بتنظيم المعلومات، بما في ذلك ما وراء البيانات، والنصوص الكاملة، وتنظيم الوسائط المتعددة.

ما وراء البيانات:

ما وراء البيانات **Metadata**، مصطلح تم سكه في نهاية تسعينيات القرن العشرين، للدلالة على تلك التوصيفات التي تعد على وجه التحديد، لتنظيم المعلومات الرقمية التي يمكن الوصول إليها على الإنترنت. وقد تم تطبيق ما وراء البيانات في البداية على المصادر الإلكترونية، والمعلومات المرتبطة بالشبكات دون غيرها. وقد اتسع مفهوم المصطلح ليشمل جميع تطبيقات تنظيم المعلومات، لأن ما وراء البيانات أصبح مصطلحا متداولا إلى حد بعيد، كما أن دور الإنترنت كمنصة للوصول إلى أنواع شتى من المعلومات في نمو مستمر.

ماذا يقصد بما وراء البيانات؟

من الممكن الآن تعريف ما وراء البيانات بطريقتين؛ أولاهما أكثر تحديدا في مجالها، وتعني التوصيفات التي تتوافر للمعلومات المرتبطة بالشبكات والمصادر الرقمية، باتباع مواصفة معيارية **standard** أو إطار **framework** (كبؤرة دبلن

Dublin Core على سبيل المثال)، يتم إعداده لهذا الغرض على وجه التحديد. والتعريف الآخر أكثر اتساعا في مجاله، ويشمل بيانات الفهرسة والتكشيف، التي يتم إعدادها لأي نوع من الوثائق، باستخدام الطرق التقليدية، من أجل وصف المعلومات وتنظيمها. وبهذا المعنى، فإن بيانات الفهرسة التي يتم إعدادها بتصنيف ديوي العشري (**DDC**) وقواعد الفهرسة الأنجلو-أمريكية / الفهرسة المقروءة آليا (**AACR/MARC**) على سبيل المثال، ينظر إليها أيضا بوصفها ما وراء البيانات.

ويمكن لما وراء البيانات أن تشتق أو تعين بواسطة المؤلفين، ومديري المستودعات، ومن يقومون بالتجهيز كطرف ثالث (**Dempsey & Heery, 1998**). ويمكن لبعض معلومات ما وراء البيانات أن تكون أيضا كامنة في بيئة التنظيم الفائق، في شكل تيجان لغة ترميز النصوص الفائقة (**HTML**) أو الروابط الفائقة. وعلى الرغم من أن ما وراء البيانات توفر أطرا لتنظيم المعلومات، وخصوصا بالنسبة للمصادر المرتبطة بالشبكات، بالإضافة إلى الأطر القائمة (كخطط التصنيف وقواعد الفهرسة مثلا)، يرى وول (**Wool (1998**) أن توفير ما وراء البيانات إنما هو في الأساس، امتداد لعملية الفهرسة التقليدية. وكما أوردنا بإيجاز في القسم ٢/ ٢/ ١، فإن عملية الفهرسة التقليدية تبدو غير مجدية بالنسبة لتنظيم المصادر الرقمية على الإنترنت، نظرا للخواص الفريدة التي نناقشها في الأقسام التالية.

خصائص المعلومات الرقمية على الإنترنت:

للمعلومات الرقمية، مقارنة بالمصادر الورقية، خصائصها التي تميزها. ويتعين على المرء الاعتماد على عتاد معين وبرمجيات معينة، للاطلاع على محتوى المعلومات الرقمية. يضاف إلى ذلك أن المعلومات الرقمية ينبغي أن تكون مسجلة بنوع من الصيغ **format** التي تتغير دائما كلما ارتفع مستوى العتاد والبرمجيات أو تطور.

وأحيانا ما تكون الطبعات المختلفة من البرمجيات نفسها متوافقة، إلا أنها لا تكون كذلك في الغالب الأعم. وتصبح قضية التوافق أسوأ حالاً، إذا كان المرء يتعامل مع معلومات رقمية ناشئة بحزم برمجيات مختلفة.

وفضلا عن ذلك، فإن المعلومات الرقمية على الإنترنت، عادة ما تسجل في شكل تنظيم فائق، يتناقض بحدة مع التنظيم المسطح للمعلومات الورقية. وييسر ظهور الإنترنت، على الإنسان التواصل وتقاسم المصادر إلا أن غياب تقنيات ضبط الجودة على الإنترنت يسفر عن فيضان من المعلومات التي تتمتع بدرجات متفاوتة من الجودة. ولهذا فإنه ينبغي اتباع طرق مختلفة لوصف وتنظيم المعلومات الرقمية على الإنترنت؛ نظرا لأن المواصفات المعيارية (كتصنيف ديوي العشرى **DDC**، وقواعد الفهرسة الأنجلو-أمريكية/ مارك / **AACR/ MARC**) قد وضعت قبل إنتاج المعلومات الرقمية بكميات ضخمة، مصممة في المقام الأول لتنظيم المعلومات الورقية. وقد ظهرت ما وراء البيانات لمجرد التعامل مع مشكلة تنظيم المعلومات الرقمية، التي تصاغ **formatted** بطريقة فريدة، وتتسم بالتغير المستمر، وتفاوت الجودة، وضخامة الكم.

أمثلة للمواصفات المعيارية لما وراء البيانات:

على الرغم من أن مفهوم ما وراء البيانات جديد نسبيا في مجال تنظيم المعلومات، فقد تم وضع عدد كبير من المواصفات المعيارية لما وراء البيانات، كما أن هناك الكثير في الطريق. وبؤرة دبلن **Dublin Core (DC)**، وإطار وصف المصادر **Resource Description Framework (RDF)**، بعض أمثلة المواصفات المعيارية لما وراء البيانات.

بؤرة دبلن:

وبؤرة دبلن **DC** أنموذج بسيط لوصف المحتوى بالنسبة للمصادر الإلكترونية **(Weibel, 1997)**. وبؤرة دبلن جهد مشترك بين خبراء في مجال المكتبات، ومجموعات البحث في المشابكة والمكتبات الرقمية، ومجموعات متعددة من اختصاصيي المحتوى، في سلسلة من ورش العمل المتفرقة، القائمة على دعوة المشاركين فيها، تسمى سلسلة ورش عمل بؤرة دبلن **Dublin Core Workshop Series**. وتتكون بؤرة دبلن من خمسة عشر عنصرا، هي العنوان، والمنشئ، والموضوع، والوصف، والناشر، والمساهم، والتاريخ، والنوع، والصيغة، ومحدد الهوية، والمصدر، واللغة، والعلاقة، والتغطية، والحقوق. وكما بين وايبل **Weibel (1997)**:

> فإن بؤرة دبلن لم يقصد بها أن تحل محل نماذج الوصف الأكثر ثراء، كالفهرسة بقواعد الفهرسة الأنجلو-أمريكية / مارك، وإنما لتوفر مجموعة بؤرية من عناصر الوصف، التي يمكن أن تستخدم من جانب المفهرسين أو غير المفهرسين، للوصف البسيط للمصادر. **(p.9)**

وتستخدم بؤرة دبلن عمليا، في المقام الأول، لتوفير توصيفات بسيطة للمصادر الشبكية، كمواقع العنكبوية على سبيل المثال. ويبدو أنها أشهر مواصفة معيارية من بين جميع نماذج ما وراء البيانات. وفضلا عن ذلك، فإن بؤرة دبلن هي الأنموذج المبدئي **prototype** الذي أعطى دفعة لتطوير إطار وصف المصادر **RDF**، في تجمع العنكبوتية العالمية **World Wide Web Consortium (W3C)**. ويرتبط القائمون على تطوير بؤرة دبلن ارتباطا وثيقا بجهد تجمع العنكبوتية العالمية، في إطار وصف المصادر.

إطار وصف المصادر:

تطور إطار وصف المصادر برعاية تجمع العنكبوتية العالمية؛ إذ أصبح إحدى توصيات هذا التجمع، في فبراير عام ١٩٩٩، ثم أصبح بعد ذلك مكونا رئيساً في العنكبوتية الدلالية، التي وضع تصورها بيرنرز-لي **Berners-Lee** والعاملون معه **(Berners-Lee, Hendler, & Lassila, 2001; Shadbolt, Hall, & Berners-Lee, 2006)**، على سبيل المثال. ويشتمل القسم ١٢/ ٣ على مناقشة للعنكبوتية الدلالية من منظور تنظيم المعلومات واسترجاعها. وكما ذكر شافر **Shafer (1998)**:

> فإن إطار وصف المصادر بنية أساس لتشفير **encoding** أو ترميز ما وراء البيانات ونمذجتها وتبادلها. وإطار وصف المصادر في جوهره، أنموذج ثلاثي الأبعاد بسيط: فما وراء البيانات تتعلق بمصدر ما، وللمصدر خاصية واحدة أو أكثر، ولكل خاصية قيمة. **(p. 21)**

ويستخدم إطار وصف المصادر لغة الترميز القابلة للتوسع **eXtensible Markup Language (XML)** بوصفها نظْم **syntax** النقل. ويتم تطوير إطار وصف المصادر لدعم الكثير من احتياجات ما وراء البيانات المختلفة، للمتعهدين وموردي المعلومات **(Weibel, 1997)**. ولا يتكفل إطار وصف المصادر بالدلالات، وإنما يكفل قدرة كل وسط أو تجمع لوصف المصادر، على تحديد معالم نظام دلالي يعبر عن متطلبات ذلك الوسط أو التجمع. وبعبارة أخرى، فإن إطار وصف المصادر، على عكس بؤرة دبلن، لا يحدد دقائق مواصفات العناصر المعينة التي ينبغي أن يتضمنها الإطار، وإنما يكفل لمن يستخدمونه حرية اختيار معالم المواصفات وتحديدها في حدود إطارهم، بناء على احتياجاتهم. ومن ثم فإن إطار وصف المصادر يشكل أساسا لتجهيز

ما وراء البيانات. وكما ذهب لاسيلا **Lassila (1997)**، فإن إطار وصف المصادر يمكن أن يستخدم في مجالات تطبيقية عدة؛ في اكتشاف المصادر، على سبيل المثال، لدعم قدرات محرك البحث، وفي الفهرسة لوصف المحتوى وعلاقات المحتوى المتاحة على موقع عنكبوتية بعينه، أو صفحة عنكبوتية بذاتها، أو مكتبة رقمية بعينها. ويشتمل القسم ١٢/ ٣/ ١ على المزيد من المعالجة لإطار وصف المصادر، في سياق النظر في الهندسة المعمارية للعنكبوتية الدلالية.

محدد هوية الكيانات الرقمية:

على عكس كل من بؤرة دبلن، وإطار وصف المصادر، فإن محدد هوية الكيانات الرقمية **Digital Object Identifier (DOI)** يوضح معالم الكيانات الرقمية كلا على حدة (كمقالات الدوريات، ومقاطع الفيديو، والمقطوعات الموسيقية على سبيل المثال) بخطته المحددة الخاصة. والمؤسسة الدولية لمحدد هوية الكيانات الرقمية **International DOI Foundation (2009)**، منظمة تأسست لتطوير وإدارة نظام محدد هوية الكيانات الرقمية، وتبين في صفحتها الرئيسة **(www.doiorg)**:

> لقد قصد بنظام محدد هوية الكيانات الرقمية **Digital Object Identifier (DOI) System**، تحديد هوية كيانات المحتوى في البيئة الإلكترونية. ويتم تخصيص أو تعيين أسماء محدد هوية الكيانات الرقمية لأي كيان، للاستعمال على الشبكات الرقمية. وتستخدم هذه الأسماء لتوفير معلومات حديثة، تشمل أماكن العثور عليها (أو على معلومات حولها) على الإنترنت. ومن الممكن للمعلومات المتعلقة بكيان رقمي ما أن تتغير من وقت لآخر، بما في ذلك مكان العثور عليه، إلا أن اسم محدد هوية الكيان الرقمي لا يتغير.

ويتكون نظام محدد هوية الكيانات الرقمية من ثلاثة أجزاء؛ هي محددات الهوية، والدليل، ومرصد بيانات نظام محدد هوية الكيانات الرقمية. ويضم كل محدد للهوية صدراً أو سابقة **prefix** وعجزا أو لاحقة أو كاسعة **suffix**، يفصل بينهما شرطة تميل لليسار **forward slash** (/). ويتم تعيين أو تحديد الصدر، الذي يعرف أيضا بمحدد هوية الناشر **publisher ID**، من قبل وكالة تسجيل محددات هوية الكيانات الرقمية للناشرين، أما العجز، الذي يسمى أيضا محدد هوية المادة **item ID**، فيعين أو يحدد من قبل الناشر **(Wang, 2007)**. وبنية محدد هوية الكيانات الرقمية وتعيينه يشبهان، في الجوهر بنية وتعيين الرقم المعياري الدولي للكتاب **(International Standard Book Number ISBN)**، والرقم المعياري الدولي للدورية **(International Standard Serial Number ISSN)**، هذان النظامان اللذان يستخدمان منذ سبعينيات القرن العشرين، للتحقق من هوية الكتب والدوريات، على نحو لا لبس فيه. يضاف إلى ذلك أن نظام محدد هوية الكيانات الرقمية يستخدم دليلا مركزيا لتدارك المشكلات المتعلقة بتغيرات ملكية المحتوى الرقمي وأماكنه. ويتم اختزان المعلومات المتعلقة بالكيان الذي تحدد هويته في مرصد بيانات نظام محدد هوية الكيانات الرقمية **(Nair & Jeavan, 2004)**.

وهناك خاصتان تميزان محددات هوية الكيانات الرقمية عن جميع طرق تنظيم المصادر الرقمية؛ أولاهما أن محدد هوية الكيان الرقمي يمكن تطبيقه للكيان الرقمي بكامله (كالدورية التخصصية على سبيل المثال) أو أي جزء منه (كالقسم، أو الشكل التوضيحي، أو الجدول على سبيل المثال) على السواء. وتكفل خاصية تطبيق محدد

هوية الكيان الرقمي هذه، بالمعنى الحرفي، تنظيم المعلومات على أي مستوى من التحليل، الأمر الذي نادراً ما يتحقق في المجال، إلا أن له مضامينه الإيجابية بالنسبة لاسترجاع الفقرات. أما الخاصية الثانية فإن محدد هوية الكيان الرقمي يتسم بالثبات والاستقرار والاطراد؛ بمجرد أن يسجل ويعين، في ربط محتواه بالدليل الذي دائما ما يشتمل على المكان الراهن لذلك الكيان الرقمي، بصرف النظر عما إذا كان سلفه يمكن أن يتغير أو يختفي من الإنترنت.

ولا يكفل نظام محدد هوية الكيان الرقمي نهجا بديلا لتنظيم المعلومات فحسب، وإنما يتيح إمكانية الوصول السلس إلى الكيانات الرقمية التي يتم التحقق من هويتها. ومن أمثلة تطبيق محددات هوية الكيانات الرقمية، منظمة كروسرف **Crossref** (**www.crossref.org**)، وهي منظمة غير ربحية، تضم دور النشر والمكتبات، تهدف إلى ربط المستفيدين، عن طريق محددات هوية الكيانات الرقمية، بالمحتوى البحثي الأولي كالكتب، ومقالات الدوريات، وأعمال المؤتمرات. ويقوم الناشرون المرتبطون بكروسرف بتعيين محددات هوية الكيانات الرقمية لكل مادة ينشرونها، ويمكن للمستفيدين بدورهم الوصول إلى تلك المواد، عن طريق روابط تحديد هوية الكيانات الرقمية.

إلا أن نظام محدد هوية الكيانات الرقمية يبدو أقرب إلى تقنية الربط الدائم الموحدة منه إلى المواصفة المعيارية لما وراء البيانات، وذلك لأن كل محدد لهوية الكيانات الرقمية يرتبط بسلسلة مما وراء البيانات، بالإضافة إلى مجموعة من

المعلومات الوراقية والتجارية المتصلة بمحتواه (كالعنوان، والمؤلف، وتاريخ النشر، والسعر، على سبيل المثال). وعندما يتم تحويل معلومات محدد هوية الكيانات الرقمية إلى الناشر، فإنه يبحث عن محدد هوية الكيانات الرقمية في بياناته الوراقية، ويعثر على المحتوى عن طريق معلومات الربط المختزنة في الدليل المركزي. ومن ثم يمكن للمحتوى الرقمي أن يظهر للمستفيد على شاشة الحاسب.

وقد تم تطوير مواصفات معيارية أخرى لما وراء البيانات، أو في سبيلها للتطوير لوصف المعلومات الرقمية على الإنترنت. وقد أجرى زنج وكين **Zeng and Qin (2008)** دراسة وصفية تحليلية شاملة للمواصفات المعيارية الحالية لما وراء البيانات، وقسما هذه المواصفات إلى ثماني فئات (المواصفات ذات الأهداف العامة، والمصادر الثقافية والبصرية، والأرشيفات والمحفوظات). وهناك تزايد في الحاجة إلى المواصفات المعيارية لما وراء البيانات هذه، وكذلك الحال أيضا بالنسبة للمخاوف والتساؤلات.

بعض التساؤلات والمخاوف حول ما وراء البيانات:

يتم إنشاء ما وراء البيانات، كما سبق أن بينا، من أجل وصف المصادر الرقمية وتنظيمها في الأساس، إلا أنه فيما يتعلق بطبيعة المعلومات الرقمية أو مصادر الإنترنت، ما جدوى ما وراء البيانات؟ ففي البيئة الورقية تتم فهرسة كل مطبوع تضمه مجموعات المكتبة. وفيما يبدو فإن الأسلوب نفسه غير قابل للتطبيق بالنسبة للمعلومات الرقمية على الإنترنت؛ نظرا للأسباب التي سبق ذكرها. ويطفو على

السطح أيضا تساؤل مماثل: من يمكن أن ينشئ ما وراء البيانات، في ظل الكميات الضخمة، والجودة غير المؤكدة للمصادر المرتبطة بالشبكات؟

ويتوقف توقع دورة حياة المعلومات الرقمية، إلى حد بعيد، على توافر التقنيات (كالعتاد والبرمجيات على سبيل المثال) التي تستخدم لأغراض إنشاء المعلومات والوصول إليها. ولما كانت تقنيات المعلومات تتطور بسرعة، فكيف يمكن أن نضمن للمصادر الرقمية، بمجرد وصفها، أن يتم العثور عليها واسترجاعها في جميع مراحل حياتها؟

وفي البيئة الورقية، يظل محتوى أي مطبوع كما هو إلى أن تصدر منه طبعة جديدة. أما في البيئة الرقمية، فإن محتوى أي وثيقة يمكن أن يتغير باستمرار. إذن، كيف نتعامل مع المحتويات الديناميكية لكثير من هذه المصادر، عندما نحاول توفير ما وراء بيانات لها؟

وكما سبق أن ناقشنا، فإن هناك مواصفات معيارية لما وراء البيانات، لتنظيم المعلومات الرقمية على الإنترنت. ولدينا في الوقت نفسه مواصفات معيارية كتصنيف ديوي العشري، وقواعد الفهرسة الأنجلو-أمريكية/ مارك، التي كنا نستخدمها على مدى سنوات، لوصف المصادر الورقية. والآن، كيف يمكن استيعاب ما وراء البيانات في الأطر القائمة الأخرى؟ وقد أشار دمبسي وهيري **Dempsey and Heary (1998)** إلى أن مجتمع المكتبات قد بدأ توا ينظر في سبل وضع ما وراء بيانات "عالم الكتاب"، وما وراء بيانات "الشبكات" في سياق الاستخدام نفسه. ومن بين مثل هذه المحاولات جهود أوسي إل سي **OCLC** في الارتقاء بتصنيف ديوي

العشري، بأدوات مثل صائغ الكلمات **wordsmith**، الذي يقتطف المفاهيم والمصطلحات الجديدة والناشئة، من نصوص غير مقيدة، ويربطها بتصنيف ديوي العشري **(Vizine-Goetz, 1997)**.

وبالإضافة إلى ما تمت معالجته هنا بإيجاز، تظل قضية القابلية للتشغيل التبادلي **interoperability** في صميم حركة ما وراء البيانات **(Rowley & Hartley, 2008)**. وتدل القابلية للتشغيل التبادلي على قدرة نظم أو مكونات متعددة، بمنصات وواجهات تعامل، وبنى بيانات مختلفة، على تبادل البيانات وتقاسمها، بأقل قدر من الفاقد في المحتوى والوظيفية **(Zeng & Qin, 2008, p. 321)**. وقد تناول زنج وتشان **Zeng and Chan (2004)** على وجه التحديد، قضايا في دعم القابلية للتشغيل التبادلي، بين نظم معالجة المعرفة، التي غالبا ما تستخدم لإنشاء ما وراء البيانات. ويبدو أنه ليس من السهل تحقيق القابلية للتشغيل التبادلي، على الرغم من بذل جهود ضخمة في ذلك الاتجاه. يضاف إلى ذلك أن لكل مواصفة من مواصفات ما وراء البيانات قضاياها ومخاوفها المرتبطة بها دون سواها؛ ولنأخذ محدد هوية الكيانات الرقمية **DOI** كمثال **(Wang, 2007)**: أي من المواصفات المعيارية لما وراء البيانات يمكن أن تستخدم عند تعيين محددات هوية الكيانات الرقمية؟ هل يمكن تعيين محددات مختلفة لهوية الكيانات الرقمية، لكل شكل ولكل طبعة؟

ولا تشمل التساؤلات والمخاوف التي أحصيناها هنا بحال، جميع التحديات التي نواجهها في تنظيم المعلومات الرقمية على الإنترنت، بنهج ما وراء البيانات. وفضلا عن ذلك، فإنه ليست هناك خطة عامة للتعامل على وجه التحديد، مع هذه التساؤلات والمخاوف. والوصف الانتقائي، والتجديد المجدول زمنياً، والحفظ الأرشيفي المخطط، من بين الممارسات الراهنة المتبعة في توفير ما وراء البيانات

للمصادر الرقمية. وما وراء البيانات، مهما كان من الممكن للممارسات الراهنة والمستقبلية أن تتغير، إنما قصد بها في النهاية تيسير تنظيم المعلومات الرقمية، بحيث يمكن أن تسترجع فيما بعد على نحو أكثر فاعلية.

النصوص الكاملة:

عندما كان حيز اختزان الحاسبات ثمينا، ووقت استخدامها باهظ التكلفة، لم يكن من الممكن سوى لمعلومات بدائل النصوص الكاملة (كالوصف الوراقي والتكشيف على سبيل المثال) أن تختزن رقميا لأغراض الاسترجاع. أما اليوم، فإنه يمكن لحاسب شخصي عادي أن يكون قويا، من حيث حيز الاختزان وطاقة المعالجة الإلكترونية. ولم يعد اختزان النصوص الكاملة واسترجاعها، في العصر الرقمي، من الممارسات النادرة.

تنظيم معلومات النصوص الكاملة:

تنظيم معلومات النصوص الكاملة التي تختزن رقميا، من نتائج تطور تقنيات المعلومات ولاشك **(Meadow, 1992)**، إلا أن تنظيم النصوص الكاملة لا هو "اعتبار كل كلمة بمثابة واصفة" **(Fugmann, 1993)** ولا هو "لا ضرورة للتكشيف". فالأمر يتطلب تنظيما خفيا **invisible** أو ضمنيا أو مطمورا **embedded**، خلاف النصوص الكاملة نفسها، لكي تكون المعلومات قابلة للاسترجاع. ويتقاسم تنظيم النصوص الكاملة معظم خواص التكشيف الاشتقاقي، بإتباع قائمة الكلمات المستبعدة **stop word list**، بالإضافة إلى رد الكلمات إلى جذورها **stemming**، والأساليب المماثلة. وما هذا النهج سوى ما أسماه لون **Luhn**

تكشيف الكلمات المفتاحية، الذي يمكن أن يتم آليا في العادة. وتستخدم نظم استرجاع الإنترنت الشهيرة، مثل جوجل **Google**، طريقة تكشيف الكلمات المفتاحية نفسها، لتنظيم النصوص الكاملة التي تجمعها في مراصد بياناتها.

صعوبات تنظيم النصوص الكاملة:

بينما يبدو تنظيم النصوص الكاملة مرغوبا فيه من جانب المستفيدين، فإن الناتج غالبا ما يكون مختلطا ومغرقاً، نظراً لمشكلة التحقيق المنخفض. ويمكن للمثال الملائم أن يكون النتائج التي تسترجعها معظم نظم استرجاع الإنترنت؛ فغالبا ما يسترجع البحث بالكلمات المفتاحية في العنكبوتية العالمية، عدة آلاف من المواقع على الأقل، لا يمكن إلا لعدد قليل منها فقط أن يكون صالحا للاستفسار الذي قدم للنظام. ويسجل فاجمان **Fugmann (1993):**

> يفرض اختزان النصوص الكاملة أعباء ضخمة، على وجه الخصوص على حيز الاختزان، وعلى وقت البحث أو التنقيب. ولا ينطوي العنصر الثاني على استنفاد وقت المعالجة بواسطة الآلات فحسب، وإنما يشمل أيضا، وهذا هو الأهم، وقت الإنسان وجلده، اللازمين لتنقية الكم الزائد من الشوشرة المسترجعة. **(p. 99)**

ولم يعد اختزان النصوص الكاملة، كما سبق أن بينا، قضية في تنظيم المعلومات واسترجاعها، وذلك بفضل تطور تقنيات استخدام الحاسبات. إلا أن جودة التنظيم بالنسبة لمعلومات النصوص الكاملة، أبعد من أن تكون مرضية، الأمر الذي يتضح دائما في النتائج التي تقدمها معظم نظم استرجاع الإنترنت. وتتوقف الاختراقات

المستقبلية في حل هذه المشكلة الجوهرية، في المقام الأول، على البحث في معالجة اللغة الطبيعية (NLP). وتشمل بعض المهام الجوهرية، التي ينبغي أن يتصدى لها الباحثون في معالجة اللغة الطبيعية: (١) إقرار البنيان النظمي **syntactic structure** للجملة (مثل: يطير الوقت كالسهم **Time flies like an arrow**) ويعرف أيضا بالإعراب، و(٢) إقرار معنى كلمة (كالواجب **duty** مثلا) في السياق، أي المعاني المتعددة لكلمة ما (**Knight, 1999**). ويحظى موضوع معالجة اللغة الطبيعية بمزيد من المناقشة في الفصل الثاني عشر "الذكاء الاصطناعي في تنظيم المعلومات واسترجاعها".

وعادة ما يكون من الصعب الوصول إلى المعلومات غير النصية كالايضاحيات والجداول، من أجل الاسترجاع في نظم معلومات النصوص الكاملة (**Fugmann, 1993**) إذا كان التنظيم يتم آليا. ولكن، ماذا عن الأنواع الأخرى من معلومات الوسائط المتعددة، كالأصوات، والصور المتحركة، التي تتاح في صيغ رقمية؟ كيف يتم تنظيم معلومات الوسائط المتعددة في نظام الاسترجاع؟ ونتناول هذا الموضوع في القسم التالي.

تنظيم معلومات الوسائط المتعددة:

هناك تزايد في معلومات الوسائط المتعددة التي يتم إنتاجها في العصر الرقمي، كما يجعل ظهور العنكبوتية العالمية الوصول إلى مثل هذه المعلومات أكثر سهولة مما كان عليه من قبل. إلا أن تدفق معلومات الوسائط المتعددة بغزارة، يمثل أيضا تحديات غير مسبوقة بالنسبة لمجال تنظيم المعلومات واسترجاعها.

أنواع معلومات الوسائط المتعددة:

الوسائط المتعددة هي أي جمع بين الصوت والصورة والمعلومات النصية، تشتمل فيه الصور على كل من الصور الثابتة والصور المتحركة. ومن الممكن معاملة كلمتي صوتي sound وسمعي audio بوصفهما مترادفين، إلا أن هناك ميلا لإحلال مصطلح "سمعي" محل "صوتي". وأحيانا ما تستعمل عبارة الوثائق الناطقة spoken documents للدلالة على المعلومات النصية المسجلة شفويا (كالأحاديث speeches and talks على سبيل المثال). وكما هو الحال بالنسبة لمعلومات الصور، تشمل الصور الثابتة كلا من اللوحات، والصور الضوئية، والملصقات وما شابهها. ويمكن للصور المتحركة أن تكون مصحوبة أو غير مصحوبة بالصوت. وعادة ما تكون الصور غير المصحوبة بالصوت رسوماً متحركة animations أو أفلاماً سينمائية صامتة. وتصبح الصور المتحركة المصحوبة بالصوت مقاطع فيديو أو أفلاماً سينمائية. ويمكن للنصوص أن تظهر في المعلومات الصوتية كتبصرات annotations، وفي الصور كعناوين أو ترجمات. ويوضح الشكل رقم ٣/ ١ الأنواع المختلفة لمعلومات الوسائط المتعددة.

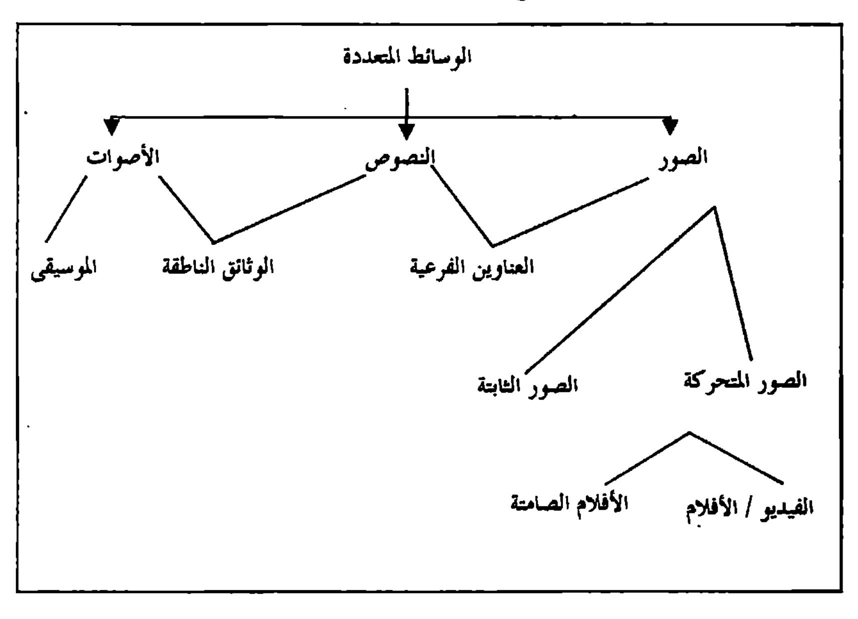
طريقتان رئيستان للتنظيم:

لقد كانت الوسائط المتعددة، فيما مضى، عادة ما تنظم بواصفات تعتمد، على سبيل المثال، على اسم المنشئ، وحجم الصورة، وعبارات أو عناوين فرعية، وكلمات مفتاحية. ومن ثم فإن هذه الطريقة تسمى الاعتماد على الوصف أو التوصيف

description based، وظلت على مدى زمني طويل تستخدم من جانب اختصاصيي المكتبات وغيرهم من اختصاصيي المعلومات، لتنظيم الوسائط المتعددة. وعلى الرغم من أن التنظيم القائم على التوصيف للوسائط المتعددة، يتعين دائما أن يتم بواسطة البشر، فإن جودة النتائج النهائية كانت تتفاوت؛ نظرا لأن الوسائط من بين أسباب أخرى، كان من الصعب في بعض الأحيان، وصفها بطريقة صريحة وموضوعية. فكيف يمكن على سبيل المثال للمرء أن يصف صراحة، صورة تشبه غروب الشمس، أو مقطوعة موسيقية، تبدو هادئة ومسالمة؟ يضاف إلى ذلك كيف يحقق البشر تنظيما مطردا دقيقا لمعلومات الوسائط المتعددة، عن طريق اتباع هذا النهج الذاتي subjective؟

وقد تطور النهج المسمى بالقائم على المحتوى **content based**، أي تنظيم الوسائط المتعددة بخواصها المميزة، كلون الصورة، وطبقة الصوت، لتجنب أوجه القصور التي تكتنف النهج القائم على التوصيف. وبناء على التقنيات الرئيسة (كالتعرف على الحديث **speech recognition**، والتعرف على الأنماط **pattern recognition**، وفهم الصور) التي تستخدم في تحليل الوسائط المتعددة لأغراض التنظيم، عادة ما يتبع النهج القائم على المحتوى، من جانب الباحثين في علوم الحاسب. كذلك يمثل هذا النهج تغيرا في الإطار النظري الأساس في تنظيم الوسائط المتعددة. وإذا ما تم تطبيق تنظيم الوسائط المتعددة القائم على التوصيف، وذلك بوصف "المضاف إليه **ofness**" (كاسم المنشئ وتاريخ النشر) و"حيثيات أو ملابسات **aboutness**" الشيء أو محتواه (كالكلمة المفتاحية والموضوع مثلا) فإن

التنظيم القائم على المحتوى يتحقق بتحليل خواص الوسائط المتعددة وصفاتها، كلون الصورة وطبقة الصوت على سبيل المثال.



الشكل رقم ٣/ ١ فئات معلومات الوسائط المتعددة

وخواص الوسائط المتعددة متعددة الأوجه أو الجوانب. وتشمل الخواص المشتركة للصور الثابتة اللون، والشكل، والقوام أو النسيج **texture**، وغير ذلك مما يمكن تحليله أيضا بالاتجاهية **directionality**، والعشوائية **randomness**، والخشونة **roughness**، والتباين **contrast**، وما شابه ذلك. وبالنسبة للمعلومات الصوتية، فإن الخواص كالطبقة والسرعة أو الإيقاع، غالبا ما توضع في الحسبان لأغراض التنظيم. ومن الممكن عادة لهذه الخصائص البدائية أو منخفضة المستوى أن تستخلص آليا، أو بطريقة نصف آلية، الأمر الذي يلغي الحاجة إلى المشاركة البشرية

باهظة التكلفة، وإن لم تكن مجزية دائما، في عملية التنظيم. ومن الممكن النزول بمستوى الخصائص بالنسبة لمعلومات الفيديو إلى تلك الخاصة بالصورة الثابتة والصوت، بعد عملية التقطيع أو التجزئ التي نتناولها في ٣/ ٣/ ٣.

ومن بين هذين النهجين المتميزين لتنظيم الوسائط المتعددة، أيهما يستخدم في الغالب الأعم في البحوث والتطبيق؟ فقد أجرت تشو **(2001) Chu** تحليلا للاستشهادات المرجعية للإنتاج الفكري المنشور في مجال تكشيف الصور واسترجاعها، وتبين لها أن النهج القائم على المحتوى كان هو الغالب في المجال، وخصوصا في السنوات الأخيرة. ويفسر التعقد المرتبط باتباع النهج القائم على التوصيف، وتطور التقنيات الرئيسة التي تستخدم في النهج القائم على المحتوى، جزئيا، نشأة ونمو الطريقة القائمة على المحتوى لتنظيم الوسائط المتعددة. إلا أن نتائج تشو لا تعني أن النهج القائم على المحتوى يمكن أن يكون الطريقة الوحيدة التي يقع عليها الاختيار لتنظيم الوسائط المتعددة. وعلى العكس، فإن النهج القائم على التوصيف، إذا ما اتبع على نحو سليم ومطرد، قادر على تحقيق تنظيم جيد، قد يكون من الصعب بلوغه بالنهج القائم على المحتوى. ويبدو أن التكامل بين النهجين المختلفين، بالجمع بين مزايا كليهما، طريق مثالي يمكن اتباعه في تنظيم الوسائط المتعددة.

التحديات في تنظيم الوسائط المتعددة:

بالإضافة إلى ما سبق أن ذكرنا من مشكلات مرتبطة بنهجي التنظيم، يبدو تجزيء **segmentation** الصور المتحركة أو تقطيعها، وتحليل الحديث **speech parsing**، حجر عثرة في تنظيم الوسائط المتعددة؛ فتجزيء الصور المتحركة، خطوة لا غنى عنها لتحليل

الصور المتحركة إلى وحدات فرعية على سبيل المثال (كعمليات الكاميرا والسكتات الملحوظة **salient stills** مثلا) والأطر الرئيسة (أي الأطر التي تمثل كل لقطة) التي يمكن حينئذ أن تكون بمثابة الأساس لتحليل المحتوى وتنظيمه **(Zhang, et al., 1995)**. ويمكن للطرق الأخرى أن تستخدم خوارزميات وأساليب مختلفة لتجزئ الصور المتحركة. إلا أن جوهر العملية يظل كما هو: تقسيم الصور المتحركة إلى كيانات أصغر لها دلالتها، بحيث يمكن تنظيمها على نحو مناسب وبدقة.

ويدل تحليل الحديث على عملية تجزيء الأحاديث الكاملة إلى فقرات وجمل وعبارات وكلمات، حتى يمكن التحقق من محتواها الموضوعي وتنظيمها. ومن الصعب صياغة المعايير الخاصة بتحليل أي من الصور المتحركة أو تجزيء الحديث؛ نظرا لطبيعة الصور المتحركة والمعلومات الصوتية؛ فالصور المتحركة متواصلة في الزمان والمكان. إلا أن الفواصل بين لقطات الكاميرا المتعاقبة لا يمكن الاعتماد عليها دائما لأغراض التجزيء. وعلى النحو نفسه أيضا، فإن معلومات الحديث لا تتضمن علامات ترقيم أو فراغات بين الكلمات، أو أي نوع آخر من العلامات، كما يمكن أن يكون عليه الحال في المعلومات التحريرية، وذلك للاسترشاد بها في التجزيء. يضاف إلى ذلك أن كثيرا من القيود التي تتصل بمعلومات الحديث، تجعل من التجزيء مهمة صعبة. ومن بين أمثلة قيود الحديث **(Spärck Jones, et al., 1996)** الأحداث أو الوقائع غير المنطوقة **nonspeech events** (كالتنفس الجهري **loud breath** وطقطقة اللسان **tongue clicks**)، ومظاهر انعدام الطلاقة **disfluencies** (كالكلمات التي لا تنطق كاملة، والوقفات، والتردد **hesitations**) والأحداث

والمكونات الوظيفية (مثل حرف العطف **and** وعبارة بالإضـافة إلى **in addition**). ونظرا لهذه المشكلات، فإن الأمر يتطلب المزيد من البحوث قبل أن تكون لدينا طـرق آلية جيدة لتجزىء الحـديث والـصور المتحركـة. ولا تـزالِ مـشاركة الإنـسان تبـدو ضرورية في تنظيم الوسائط المتعددة، حتى باتباع النهج القائم على المحتوى.

ويمكن القول بوجه عام، أن مـا أجـري مـن بحـوث حـول تنظـيم الوسـائط المتعددة أقل مما أجري حول تنظيم المعلومات النصية. ومن بين جميع أنـواع الوسـائط المتعددة المختلفة، تبدو المعلومات الصوتية أقل حظا في الدراسة مـن حيـث التنظـيم. وفي الوقت نفسه، يشهد العـصر الرقمـي كميـات متزايـدة مـن معلومـات الوسـائط المتعددة. ومن ثم، فإن سـبل تنظـيم معلومـات الوسـائط المتعـددة تفـرض تحـديات ضخمة، إذا ما أردنا أن نجعل معلومات الوسائط المتعددة من الممكن الوصـول إليهـا عند الحاجة.

مزيد من الاستطراد حول تنظيم المعلومات:

تنظيم المعلومات أمر جوهري لا غنى عنه، في اسـترجاع المعلومـات لـسببين: أولهما أن المعلومات ينبغي أن تنظم لكي يكون مـن الممكـن اسـترجاعها، وثـانيهما أن جودة التنظيم تـؤثر في أداء الاسـترجاع عـلى نحـو مبـاشر. يـستند جهـدنا في تنظـيم المعلومـات، بوجـه عـام، إلى سـمات **ofness** المعلومـات، وحيثياتهـا أو محتواهـا **aboutness**، فيما عدا النهج القائم على المحتوى بالنسبة لتنظـيم الوسـائط المتعـددة.

وتشمل السمات خواص كالمؤلف واللغة وتاريخ النشر، بينما تتعلق الحيثيات بالمحتوى الموضوعي للمعلومات. وللمقارنة فإن تنظيم المعلومات القائم على السمات الشكلية أو الخارجية أو الملامح، مباشر واضح المعالم وأقل إثارة للتحديات، بينما طرق تنظيم الحيثيات أو المحتوى أكثر صعوبة في تطبيقها.

ويعاني أداء الاسترجاع، ما لم يتم تنظيم المعلومات على نحو سليم مناسب. ولا يمكن المبالغة في تأكيد أهمية أن يكون لدينا تنظيم جيد للمعلومات. ومن ثم فإننا ينبغي أن نرمي إلى التنظيم الجيد للمعلومات، حتى يكون بإمكاننا العثور على ما نبحث عنه في عالم المعلومات الرقمية.

المراجع

Berners-Lee, Tim, Hendler, James, and Lassila, Ora. (2001). The semantic web. Scientific American, 284(5), 35-43.

Chu, Heting. (2001). Research in image indexing and retrieval as reflected in the literature. Journal of the American Society for Information Science and Technology, 52(2), 1011-1018.

Dempsey, Lorcan, and Heery, Rachel. (1998). Metadata: A current view of practice and issues. Journal of Documentation, 54(2), 145-172.

Fugmann, Robert. (1993). Subject analysis and indexing: theoretical foundation and practical advice. Frankfurt/Main, Germany: Indeks Verlag.

International DOI Foundation. (2009). Welcome to the DOI system. Retrieved July 18, 2009, from www.doi.org

Knight, Kevin. (1999). Mining online text. Communications of the ACM, 42(11), 58-61.

Lassila, Ora. (1997). Introduction to RDF metadata: W3C note 1997-11-13. Retrieved January 19, 2009, from www.w3.org/TR/NOTE-rdf-simple-intro-971113.html

Meadow, Charles T. (1992). Text information retrieval systems. San Diego, CA: Academic Press.

Nair, Saji S., and Jeevan, V.K.J. (2004). A brief overview of metadata formats. DESIDOC Bulletin of Information Technology, 24(4), 3-11.

Roweley, Jennifer E., and Hartley, Richard J. (2008). Organizing knowledge: An introduction to managing access to information. Burlington, VT: Ashgate.

Shadbolt, Nigel, Hall, Wendy, and Berners-Lee, Tim. (2006). The semantic web revisited. IEEE Intelligent Systems, 21(3), 96-101.

Shafer, Keith. (November/December 1998). Mantis Project Provides a toolkit for cataloging. OCLC Newsletter, 21-23.

Spärck Jones, Karen et al. (1996). Experiments in spoken document retrieval. Information Processing & Management, 32 (2) 399-414.

Vizine-Goetz, Diane. (1997). From book classification to knowledge organization: Improving internet resource description and discovery. Bulletin of the American Society for Information Science, 24(1), 24-26.

Wang, Jue. (2007). Digital Object Identifiers and their use in libraries. Serials Review, 33(3), 161-164.

Weibel, Stuart. (1997). The Dublin Core: A simple content description model for electronic resources. Bulletin of the American Society for Information Science, 24(1), 9-11.

Wool, Gregory. (1998). A meditation on metadata. Serials Librarian, 33(1/2), 167-178.

Zeng, Marcia L., and Chan, Lois Mai. (2004). Trends and issues in establishing interoperability among knowledge organization systems. Journal of the American Society for Information Science and Technology, 55(5), 377-395.

Zeng, Marcia Lei, and Qin, Jian. (2008). Metadata. New York: Neal-Schuman Publishers.

Zhang, H.J., et al. (1995). Video parsing, retrieval and browsing: An integrated and content-based solution. Proceedings of ACM Multimedia (pp. 15-24), New York: ACM Press.

الفصل الرابع

اللغة في تنظيم المعلومات واسترجاعها

اللغة كما أشرنا بإيجاز في ١/ ٣/ ٣، مكون أساس في كل من تنظيم المعلومات واسترجاع المعلومات (في صياغة الاستفسارات على سبيل المثال). وتتخذ اللغة في تنظيم المعلومات واسترجاعها، شكل اللغة الطبيعية أو اللغة المقيدة. وحيثما يكون هناك خيار، فإن التساؤل يثار حول أي نوع من اللغة.يمكن للمرء أن يستخدم لتنظيم المعلومات واسترجاعها.

اللغة الطبيعية:

اللغة الطبيعية هي اللغة التي يتحدث بها ويكتب البشر. وفي اللغة الطبيعية لا يُبذل أي جهد في تنظيم المعلومات واسترجاعها، للحد أو تحديد معالم المفردات vocabulary، والنظم syntax، والدلالة semantics، والعلاقات المتبادلة فيما بين المصطلحات. ويسمى ما يستخدمه البشر من أجل تنظيم المعلومات أو صياغة الاستفسارات دون اللجوء إلى لغة مقيدة باللغة الطبيعية.

وعادة ما تستعمل اللغة الطبيعية بثلاث طرق، لتنظيم المعلومات واسترجاعها. والطريقة الأولى هي استعمال المصطلحات الواردة في العناوين (ككشاف الكلمات المفتاحية في السياق Keyword In Context KWIC، على سبيل المثال)، والجمل الموضوعية أو عناوين الفقرات topic sentences، والمكونات المهمة الأخرى للوثيقة، لتنظيم المعلومات. والطريقة الثانية هي استعمال المصطلحات المستقاة أو

المشتقة من أي جزء بالوثيقة، لتنظيم المعلومات (كالتكشيف الاشتقاقي مثلا). أما الطريقة الثالثة فهي استعمال كلمات أو عبارات مقتطفة مباشرة من أسئلة المستفيدين، للتعبير عن الاستفسار، وهذا موضوع نناقشه في الفصل الخامس.

وتتكون اللغة الطبيعية أساسا، من نوعين من الكلمات؛ هما الكلمات المهمة والكلمات الوظيفية. والكلمات المهمة هي تلك المصطلحات التي تحمل معانٍ جوهرية أو مفاهيم موضوعية، بينما تشمل الكلمات الوظيفية الأدوات، وحروف الجر والروابط conjunctions، مثل(*) an, a, the, and, for, of, to, this, that, her, their. وعندما تستخدم اللغة الطبيعية في تنظيم المعلومات واسترجاعها آليا، عادة ما يتم إعداد قائمة بالكلمات المستبعدة stop-word list (أو قائمة استبعاد stop list للإيجاز)، وتختزن في النظام لتحول دون أن تصبح تلك الكلمات مصطلحات تكشيف أو مصطلحات استفسار. وتضم قائمة الاستبعاد جميع الكلمات الوظيفية المحتملة، بالإضافة إلى تلك المصطلحات المغرقة في العمومية أو المألوفة، على النحو الذي يحول دون صلاحيتها كمداخل للتنظيم. وأحيانا ما تسمى مثل هذه المصطلحات بالكلمات المهنية trade؛ فكلمة الهندسة وحدها على سبيل المثال، قد تكون لا معنى لها كمصطلح كشفي أو مصطلح استفسار، في مرصد بيانات هندسي. وتدخل المصطلحات ذات الطابع المؤقت ephemeral، التي تعرف بالتعبيرات الطنانة buzzwords، في قائمة الاستبعاد أيضا، نظرا لأنها قد لا تصمد طويلا كمصطلحات تكشيف أو استفسار مرشحة. وتظهر معظم الكلمات الطنانة بوصفها تتصل بالتقنيات، على غرار الجوفر gopher، وإنكار الخدمات(**) denial of service. ولا يمكن لأي من الكلمات التي ترد بقائمة الاستبعاد، أن تستعمل لأغراض التنظيم أو

(*) أو ما يقابل ذلك في أي لغة بالطبع. (المترجم)

(**) إحدى جرائم الإنترنت. راجع: إيان جيه. لويد. تشريعات تقنيات المعلومات. ط٥. ترجمة حشمت قاسم. القاهرة، المركز القومي للترجمة، قيد الطبع. (المترجم)

الاسترجاع. إلا أن كل نظام لاسترجاع المعلومات ينبغي أن تكون له قائمة الاستبعاد الخاصة به، التي يتم إعدادها خصيصاً للنظام، من خلال النظر في المستفيد المستهدف وعوامل أخرى.

وفي مقابل قائمة الاستبعاد، يمكن أيضا إعداد قائمة اعتبار go list لنظم استرجاع اللغة الطبيعية (Rowley, 1992). وجميع الكلمات المهمة، فيما عدا المصطلحات المهنية والتعبيرات الطنانة، مرشحة لعضوية قائمة الاعتبار. وكقائمة الاستبعاد، فإن قائمة الاعتبار قائمة للقراءة بواسطة الآلات، ومن الممكن النظر فيها أثناء التنظيم والاسترجاع. وتبدو قائمة الاعتبار في اللغة الطبيعية بوجه عام، أقل استعمالا في الغالب، من قائمة الاستبعاد، لسببين؛ أولهما أن قائمة الاستبعاد أقصر وأيسر في تداولها، ويمكن لهذه التقنية أن تختبر بوجه عام، استعمال اللغة الطبيعية في تنظيم المعلومات واسترجاعها، وثانيهما أن قوائم الاعتبار تستعمل في غالب الأحيان، في نظم اللغات المقيدة، في شكل مكانز، وقوائم رؤوس موضوعات، وخطط تصنيف.

وقد تزايد في السنوات الأخيرة استعمال قوائم الكلمات، بوصفها لغة شبه منضبطة، في نظم الاسترجاع الآلية. ونظرا لأن قوائم الكلمات غالبا ما تشتمل على مترادفات وأضداد للمصطلحات التي ترد في الوثائق، قلما تنطوي على ضبط آخر للغة، فإنها تبدو شكلا آخر لقائمة الاعتبار. وتنشئ بعض نظم استرجاع الإنترنت، وتحرص على صيانة قوائم كلمات أو قوائم اعتبار لاستعمالها الخاص.

اللغة المقيدة:

اللغة المقيدة controlled vocabulary لغة اصطناعية، حيث إن مفرداتها ونظمها syntax، ودلالتها وسياقاتها محدودة ومحددة المعالم (Wellisch, 1995). ومن الممكن للكلمة نفسها في لغات مقيدة مختلفة، أن تكون لها معان مختلفة تبعا لتوجه

اللغة. وينبغي أن يتم بناء اللغة المقيدة وصيانتها، واضعين في الحسبان مجالا موضوعيا معينا أو عدة مجالات، ويتم انتقاء المصطلحات التي تدخل في اللغة المقيدة، باتباع مبدأ مسوغ الإنتاج الفكري أو المسوغ الأدبي literary warrant، أو مسوغ المستفيد user warrant. ويعني المسوغ الأدبي أن يتم اختيار المصطلحات التي تتضمنها اللغة المقيدة، من الإنتاج الفكري القائم. وعلى النحو نفسه، فإن مسوغ المستفيد يعني أن تكون المصطلحات التي يتم اختيارها لتتضمنها اللغة المقيدة، قد استعملت بالضرورة في الماضي. والمكانز، وقوائم رؤوس الموضوعات، وخطط التصنيف هي الأنواع الثلاثة الرئيسة للغات المقيدة.

المكانز:

المكنز thesaurus، لغة مقيدة للمصطلحات باللغة الطبيعية، يتم إعداده للربط اللاحق (National Information Standard Organization, post-coordination, 1993). كذلك تعرف راولي (1992) Rowley المكنز على النحو التالي:

> تجميع للكلمات والعبارات، يبرز المترادفات وعلاقات التقسيم الهرمي، وغيرها من العلاقات والارتباطات، ومهمته توفير لغة معيارية لنظم اختزان المعلومات واسترجاعها. (P. 252)

وينطوي الربط اللاحق على جهد بشري تحكمي manipulative، وهو أحدث بكثير من لغات الربط المسبق pre-coordination، كخطط التصنيف. و تكفل لغات الربط اللاحق، كالمكانز مثلا، للمستفيدين القدرة على الربط بين المصطلحات أثناء التنظيم والاسترجاع. ووجه القصور الرئيس في الربط اللاحق هو الربط المزيف false coordination. وكمثال على ذلك فإن كلمتي "حاسب" و"مكتب" يمكن أن يرتبطا معا بشكلين؛ "مكتب الحاسب" و"حاسب المكتب"، تبعا للمقصد الأصلي؛

فإذا كان المقصود هو "حاسب المكتب" فإن نتائج الربط مثل "مكتب الحاسب" يمكن النظر إليها بوصفها ربطا مزيفا.

وتستخدم الرموز notations المعيارية في المكانز، لتحديد العلاقات التفريعية، والترابطية، والقائمة على الجناس أو المشترك اللفظي، وغيرها مما بين المصطلحات الواردة من علاقات. ويدل المختصران USE وUF (يستعمل لـ) على أي مصطلحات المكنز تعد واصفات مفضلة، ويحدد المختصر SN (بيان المجال scope note) معالم معاني الواصفات. ويدل NT (مصطلح أضيق narrower term) وBT (مصطلح أعرض broader term) على العلاقات الهرمية فيما بين الواصفات. ويتم التعبير عن علاقات التداعي أو الترابط associative بين الواصفات بالرمز RT (مصطلحات متصلة أو مرتبطة related terms). وعادة ما ترتب المكانز هجائيا وهرميا. أما شكل العرض أو الترتيب الدوار أو التبادلي، أو النسقي أو التصنيفي، والتصويري graphic، فأحيانا ما نجده في مكانز معينة أيضا (Aitchison, Gilchrist & Bawden, 1997).

والمكانز هي أكثر اللغات المقيدة المستعملة شيوعا، في تنظيم واسترجاع المعلومات غير الكتب؛ نظرا لدقتها ومرونتها وقدرتها على التعامل مع المفاهيم المركبة. وواصفات مكنز مركز معلومات المصادر التربوية Educational Resources Information Center (ERIC)، مثال على ذلك.

قوائم رؤوس الموضوعات:

قائمة رؤوس الموضوعات، لغة مقيدة للمصطلحات في اللغة الطبيعية، يتم إعدادها لكل من الربط المسبق والربط اللاحق. ويجمع الربط المسبق الذي كان سائدا قبل أربعينيات القرن العشرين بين المصطلحات قبل استعمالها لأغراض التنظيم والاسترجاع. فإذا كان المرء ينوي، على سبيل المثال، تنظيم أو استرجاع وثائق حول

"نظم استرجاع الإنترنت internet retrieval systems"، فإن العبارة المكونة من ثلاث كلمات، يمكن أن تظهر على وجه التحديد في صيغة مترابطة سلفاً، ومن ثم فإن اللغات المقيدة المترابطة مسبقا أو سلفاً، ليست تحكمية nonmaipulative. ولما كانت قوائم رؤوس الموضوعات تكفل كلا من الربط المسبق والربط اللاحق، فإنها تبدو أكثر مرونة من خطط التصنيف، ولكنها أقل مرونة من المكانز.

وتسمى المصطلحات، في قوائم رؤوس الموضوعات برؤوس الموضوعات، وترتب هجائيا. وتشمل الرموز بالنسبة لرؤوس الموضوعات، انظر see، للإحالة من المصطلحات غير المُفضلة إلى المفضلة، في حين يمثل X الذي يعني "انظر من" *See from* التعبير المقابل لإحالة "انظر"، كما في "معاق handicapped" انظر "عاجز بدنيا physically challenged"، و"عاجز بدنيا" انظر "معاق". وهناك مجموعة أخرى من الرموز الخاصة بقوائم رؤوس الموضوعات، تشمل "انظر أيضا see also"، وXX الذي يفيد "انظر أيضا من see also from"، وتدل إحالة "انظر أيضا see also" على كل من العلاقات التفريعية والترابطية بين رؤوس الموضوعات التي يقع عليها الاختيار. إلا أن ذلك الأسلوب ينال من مدى تخصيص قوائم رؤوس الموضوعات ودقتها. وعلى غرار الرمز X، يدل XX على التعبير التقابلي لعبارة انظر أيضا.

وتستعمل قوائم رؤوس الموضوعات، بوجه عام، لكل من تنظيم المعلومات واسترجاعها، على الرغم من أنه قد يبدو أنها تستخدم على نحو أقل من المكانز، لأغراض تنظيم واسترجاع المعلومات غير الكتب. وقائمة سيرز لرؤوس الموضوعات *Sears List of Subject Headings* من أمثلة هذه القوائم. والمثال الآخر الشهير هو رؤوس موضوعات مكتبة الكونجرس *Library of Congress Subject Headings*، التي تبدو رموزها في الطبعة الحادية عشرة أقرب شبها بالمكانز.

خطط التصنيف:

خطة التصنيف، لغة مقيدة من مصطلحات في شكل هجائي رقمي، توضع في المقام الأول للربط المسبق. وتسمى الوحدات في خطط التصنيف بالفئات classes، التي تميز بعد ذلك رقميا، أو هجائيا، أو هجائيا رقميا. وبوصفها أول أنواع اللغات المقيدة التي تم وضعها، شهدت خطط التصنيف، وعايشت تغيرات عدة، طوال القرون الماضية. ومن ناحية أخرى يتم تحديث هذه الخطط ومراجعتها، مراراً وتكراراً بمرور الوقت.

وعلى عكس المكانز، فإن قوائم رؤوس الموضوعات وخطط التصنيف، يتم بناؤها وفقا لإطار اصطناعي للمعرفة. فقد جنح تصنيف ديوي العشري، على سبيل المثال، لتقديم عالم المعرفة على أساس عشرة أقسام. ويتم تشكيل عشر فئات على وجه التحديد، عند كل مستوى للتفريع في إطار الخطة، الأمر الذي يدل على جمود النظام. ويتم إبراز العلاقة التفريعية بين الفئات باستعمال مسافات أو أبعاد مختلفة لرموز التصنيف؛ أي إنه كلما طال رمز التصنيف، ازداد عمق مكان الفئة في البنيان الهرمي. ويتم إبراز العلاقات الترابطية بين المصطلحات عن طريق أسلوب "انظر" و"انظر أيضا" الذي لا يستخدم إلا عرضا.

وغالبا ما يتم اختيار خطط التصنيف تقليديا، لتنظيم واسترجاع معلومات الكتب. وتصنيف ديوي العشري، وتصنيف مكتبة الكونجرس مثالان لخطط التصنيف.

المكانز في مقابل قوائم رؤوس الموضوعات وخطط التصنيف:

يلخص الجدول رقم ٤/ ١ خصائص كل من هذه الأنواع الثلاثة للغات المقيدة. وبالإضافة إلى ما تمت مناقشته، فإن طريقة التحليل خاصية أخرى ترتبط باللغات المقيدة. ويعني الحصر enumeration بوصفه طريقة للتحليل، ببساطة، توفير

قائمة بالمصطلحات دون القدرة على الربط فيما بينها، للتعبير عن شيء ما أكثر تعقدا. أما التركيب synthesis فيكفل، على العكس، الربط بين بعض المصطلحات، لتكوين مصطلحات أكثر تعقدا، في مرحلة التنظيم أو المدخلات، أو في مرحلة إجراء عمليات البحث والاسترجاع (Lancaster, 1986). وترتبط طريقة التحليل ارتباطا وثيقا بطريقة الربط. أي أن لغة الربط المسبق حصرية، بينما لغة الربط اللاحق تحليلية تركيبية، كذلك تحدد طرق التحليل والربط مدى تخصيص اللغة المقيدة ومرونتها. وتبدو اللغة التحليلية التركيبية ولغة الربط اللاحق أكثر تحديدا أو تخصيصا وقابلية للاستيعاب من اللغة الحصرية أو لغة الربط المسبق.

ووفقا للجدول رقم ٤ / ١ تبدو المكانز أكثر تخصيصا ومرونة من خطط التصنيف، بينما تقع قوائم رؤوس الموضوعات في مرتبة وسط بينهما على هذا الطيف، الأمر الذي يفسر السبب في أن تصبح المكانز أوسع اللغات المقيدة انتشارا في الاستخدام.

اللغة الطبيعية في مقابل اللغة المقيدة:

ناقش القسمان السابقان ملامح وخصائص كل من اللغة الطبيعية واللغة المقيدة. وآن الأوان للنظر في أوجه الاختلاف فيما بينهما، والقاسم المشترك بينهما كلغات لتنظيم المعلومات واسترجاعها.

الحقب المختلفة للغات تنظيم المعلومات واسترجاعها:

تعد اللغات المقيدة، في تاريخ تنظيم المعلومات واسترجاعها، وافدا جديدا بالمقارنة باللغة الطبيعية. ومن الممكن التحقق من معالم أربع حقب في تطور لغات تنظيم المعلومات واسترجاعها حتى الآن.

وتدل الحقبة الأولى على زمن ما قبل ظهور أي لغة مقيدة. فقد كانت اللغة الطبيعية هي اللغة الوحيدة المستخدمة، بينما كان تنظيم المعلومات واسترجاعها يتشكل كمجال موضوعي. وبدأ الناس يدركون أوجه القصور (كالمشترك اللفظي والمترادفات مثلا) في اللغة الطبيعية، عندما كانت تستعمل لتنظيم المعلومات واسترجاعها.

الجدول رقم ٤/ ١ مقارنة اللغات المقيدة

الخصائص / اللغة المقيدة	المكانز	قوائم رؤوس الموضوعات	خطط التصنيف
مكونات المصطلح	الواصفات	رؤوس الموضوعات	وسيمات التصنيف
رموز الإحالات	U, UF, SN, BT, NT, RT	انظر، انظر أيضا X, XX	انظر، انظر أيضا
طريقة التحليل	التحليل والتركيب ↔ الحصر		
طريقة الربط	الربط اللاحق ↔ الربط المسبق		
التخصيص	أكثر ↔ أقل		
المرونة	أكثر ↔ أقل		
نوعية الوثائق المستهدفة	غير الكتب	الكتب وغير الكتب	الكتب

وتدل الحقبة الثانية على دخول اللغة المقيدة إلى تنظيم المعلومات و استرجاعها. ففي البداية كانت نظم الربط المسبق كخطط التصنيف، تستخدم بكثافة. ثم ظهرت أنواع أخرى من اللغات المقيدة، وهي قوائم رؤوس الموضوعات والمكانز فيها بعد. وقد تعايش كلا النوعين من لغات تنظيم المعلومات واسترجاعها، في هذه الحقبة، مع الميل نحو اللغة المقيدة. كذلك بدأ الجدل حول اللغة الطبيعية في مقابل اللغة المقيدة حينئذ، ولا يزال متواصلا.

وتتميز الحقبة الثالثة بنهضة اللغة الطبيعية من جديد، نتيجة لأساليب استرجاع الكلمات المفتاحية، وتطور نظم النصوص الكاملة. ولا تزال اللغة المقيدة تستخدم لتنظيم المعلومات واسترجاعها، ولكن في نظم الاسترجاع الوراقية (الببليوجرافية) في الغالب الأعم. فهل يمكن للغة المقيدة أن تظل لغة صالحة لتنظيم المعلومات واسترجاعها؟ فقد ثار الكثير من الجدل في تلك الحقبة فيما يتعلق باللغة الطبيعية في مقابل اللغة المقيدة.

ثم بدأت الحقبة الرابعة، عندما بدأ استعمال واجهات التعامل باللغة الطبيعية في تنظيم المعلومات واسترجاعها، واستخدام تقنية لغة مقيدة خلف ستار. ونظرا لأن نهج اللغة المقيدة لا يمكن رؤيته من جانب المستفيد النهائي، فقد أسمتها ميلستيد Milstead (1995) اللغة المقيدة الخفية في بيئة استرجاع اللغة الطبيعية. ومن بين الأمثلة على ذلك إخفاء اللغة المقيدة وراء غلاف shell اللغة الطبيعية. وقد أسهمت مظاهر التقدم في معالجة اللغة الطبيعية، على نحو ملحوظ أيضا في هذا التطور. وقد بدأت بعض نظم استرجاع المعلومات، كنظام وست المسمى ون West's WIN (وستلو طبيعي Westlaw is Natural) تقدم خدمات استرجاع باللغة الطبيعية.

وعلى الرغم من أن الحدود الفاصلة بين مختلف الحقب، لا تبدو على قدر كبير من الوضوح والتحديد، فإن لغة تنظيم المعلومات واسترجاعها، قد اجتازت بجلاء الحقبتين الأوليين، وتمر الآن بالحقبتين الثالثة والرابعة.

لماذا اللغة الطبيعية أم لماذا اللغة المقيدة؟

ينبغي أن تكون هناك أسباب وراء تعايش اللغة الطبيعية مع اللغة المقيدة، في تنظيم المعلومات واسترجاعها. ولا تخرج الأسباب عن وجهات النظر المؤيدة ووجهات النظر المعارضة لكلا النوعين من لغات تنظيم المعلومات واسترجاعها.

قضية الترادف:

قضية المترادفات إحدى أكثر الحجج التي تساق ضد استعمال اللغة الطبيعية، في تنظيم المعلومات واسترجاعها؛ فالمترادفات في اللغة الطبيعية مصطلحات مختلفة تؤدي المعنى نفسه. فالحاسبات المصغرة microcomputers، والحاسبات الشخصية personal computers، وبي سي PCs، وسطح المكتب desktop، والمفكرات notebooks، وسطح الركبة laptop، وما هو على شاكلتها مترادفات(*). إلا أن أي مصطلح في هذه القائمة الجزئية يمكن أن يقع عليه الاختيار، ويستعمل لأغراض تنظيم المعلومات واسترجاعها؟ وهذه إحدى قضايا الجدل حول لغات تنظيم المعلومات واسترجاعها. وعندما تكون اللغة المقيدة هي لغة تنظيم المعلومات واسترجاعها، التي وقع عليها الاختيار، فإن المترادفات لا تصبح مصدرا للمخاوف، لأن مصطلحا واحدا، وواحد فقط هو ما يقع عليه الاختيار في اللغة المقيدة بوصفه المصطلح المفضل. أما بقية المترادفات فتعامل بوصفها مصطلحات غير مفضلة، ويحال منها إلى المصطلح المفضل.

(*) مثال غير موفق للأسف، لأن البي سي اختصار مباشر للحاسب الشخصي، والمفكرات وسطح الركبة وغيرها، من الحاسبات المحمولة، وإن كانت تستخدم لأغراض شخصية، تختلف عن الشخصي وسطح المكتب، ومن ثم فإنه يمكن النظر إلى هذه المصطلحات بوصفها أشباه مترادفات. (المترجم)

قضية المشترك اللفظي:

تبدو الكلمات المتفقة شكلا المختلفة معنى homographs، الناتجة عن الظاهرة المعروفة بتعدد المعاني polysemy، حجة أخرى، غالبا ما تساق في الجدل حول اللغة الطبيعية في مقابل اللغة المقيدة. ويعني المشترك اللفظي أو الجناس، وجود مصطلحات هجاؤها واحد ولكنها تحمل معانٍ مختلفة في السياقات المختلفة. ومصطلحات مثل record، وsubject، وdrug، وspring، وbank، وduty مجرد أمثلة قليلة للمشترك اللفظي(*). وفي اللغة الطبيعية في تنظيم المعلومات واسترجاعها، يمكن للمشترك اللفظي، نظرا لغياب السياق، أن يسفر عن مشكلات كالغموض مثلا. وقد تبين للانكستر وورنر Lancaster and Warner (1993) أن مشكلة الغموض في الاسترجاع، الناتجة عن المشترك اللفظي غالبا، ما تكون نظرية أكثر منها فعلية، نظرا لأن الكلمات التي يمكن أن تكون غامضة في حد ذاتها، لا تصبح كذلك عندما تستعمل مرتبطة بكلمات أخرى(**). حقا يمكن للمشترك اللفظي ألا يصبح غامضا في معناه، إذا ما استعمل مرتبطا بكلمات أخرى، إلا أن الارتباط في الاستعمال قد يحدث أو لا يحدث في اللغة الطبيعية لتنظيم المعلومات واسترجاعها. وفي مقابل ذلك، فإن السياق اللازم لتفسير المشترك اللفظي، دائما ما يتوافر عن طريق وسائل مختلفة في اللغة المقيدة. ومن بين الوسائل المألوفة استعمال الأقواس لتحديد سياق المشترك اللفظي، كما في duty (المسئولية) وduty (الضرائب)(***).

قضية النَظْم(****):

للغة نَظْمها، ولكن كيف يمكن التعبير عن النَظْم syntax عندما يقع الاختيار على اللغة الطبيعية لتنظيم المعلومات واسترجاعها؟ هب أن وثيقة ما تم تمثيلها

(*) يناظر ذلك في العربية "عين" و"قرار" على سبيل المثال. (المترجم).
(**) أي عندما توضع في سياق معين. (المترجم)
(***) يقصد بذلك تحديد المجال الموضوعي لاستعمال الكلمة. (المترجم)
(****) النحو **grammar** مستويان؛ الصرف **morphology** أي بناء الكلمة، والنظم **syntax** أي بناء العبارة أو الجملة. وهناك من يستعمل "النحو" مقابلا لكلمة "**syntax**"، إلا أننا نرى في ذلك ما يجافي مقتضيات الدقة. (المترجم)

بالمـصطلحات الثلاثـة "الولايـات المتحـدة الأمريكيـة *USA*" و"الـسيارات *automobiles*" و"اليابان *Japan*"، في اللغة الطبيعية؛ فهذه الوثيقة يمكن أن تـدور على سبيل المثال، حول تصدير اليابان للسيارات إلى الولايات المتحـدة الأمريكيـة، أو العكس، تصدير الولايات المتحدة للسيارات إلى اليابان. ومن غير الواضـح أي دولـة هي المصدِّر، عندما ترد هذه المصطلحات الثلاثة فقط دون أي معلومات عن النَظْم(*). ومن الممكن حل هذه المشكلة بسهولة في اللغة المقيدة، بوسائل كالأدوار roles، وهي رموز أو أرقام تدل على العلاقة النظمية بين المصطلحات. ومن الممكن في هـذا المثـال تعريف الرقم ١ بوصفه دور المصدِّر، ثم نضعه بعد المـصطلح "اليابـان *Japan*"، أي (1) Japan. وعلى النحو نفسه يمكن تسمية الرقم ٢ بوصفه دور المستورد، ثـم نعينـه للمصطلح *USA*، أي *(2)* USA. ومن شأن مثل هذه الطريقة المتبعة في اللغة المقيـدة، أن تكفل لنا التغلب على قضية النَظْم، بينما لا تكفل اللغة الطبيعية ذلك.

قضية الدقة:

دائما مـا يكـون مـن المرغـوب فيـه، أن تكـون هنـاك لغـة لتنظـيم المعلومـات واسترجاعها، يمكن أن تعبر بدقة عن المعلومات واستفسارات البحـث. ويبـدو هـذا الهدف قابلا للتحقيق إذا ما وقع الاختيار على اللغـة الطبيعيـة، بوصـفها لغـة لتنظـيم المعلومات واسترجاعها. والمبرر المنطقي لهذه العبارة ثنائي الأبعاد؛ أولا، لا حاجة إلى معالجة إضافية (كالشرح مثلا) إذا ما استعملت اللغة الطبيعية للتنظيم والاسـترجاع، ثانيا لا ضرورة للتفسير عند استعمال اللغة الطبيعية في التنظيم والاسـترجاع، لأن مـا يـستعمله المؤلـف أو المـستفيد إنـما هـي اللغـة التـي تـستعمل في تنظـيم المعلومـات

(*) ربما يدرك الناطقون بالعربية مدى عبقرية نظام النحو العربي المرتبط بالدلالة؛ إذ لا يمكن معرفة معنى الكلمة على وجه التحديد إلا بإعرابها أي بيان موقعها في الجملة. والقراءة السليمة مرتبطة بالإعراب. وربما يدل ذلك على مدى ما فقدت اللغة العربية نتيجة لتجاهل الحركـات في معظـم النصوص المطبوعة. (المترجم)

واسترجاعها. وبالمقارنة، فإن اللغة المقيدة اصطناعية، ولا تدل على ثراء اللغة الطبيعية، التي تستعمل في إنشاء الوثائق والتعبير عن استفسارات البحث. كذلك تفتقر اللغة المقيدة أيضا إلى التحديد أو الدقة، نتيجة لما تتعرض له في عملية المعالجة. ويبدو تفسير اللغة المقيدة أمرا لا غنى عنه؛ نظرا لأن مضمون كل مصطلح ودلالته إنما تتحدد معالمها بوضع متلق مستهدف بعينه في الحسبان. ويمكن لمثل هذا التفسير أن يفضي حتما إلى عدم الدقة في لغة التنظيم والاسترجاع التي تعتمد على اللغة المقيدة(*).

قضية التحديث:

لا تحتاج اللغة الطبيعية إلى أي تحديث updating، بينما تحتاج اللغة المقيدة إلى ذلك؛ فمن الممكن للمصطلحات الجديدة أن تستعمل لأغراض تنظيم المعلومات واسترجاعها بمجرد أن ترد في اللغة الطبيعية، بينما لا يمكن حدوث أي تغيير في اللغة المقيدة، إلا حين تعرضها لعملية التحديث الجامدة المطولة. ومن ثم فإن المصطلحات في اللغة المقيدة ليست حديثة على الدوام. ولا يمكن بحال الرد على الاستفسارات التي تتضمن مصطلحات جديدة، إذا ما وقع الاختيار على اللغة المقيدة للتنظيم والاسترجاع(**).

(*) هناك من يرى أن تدخل اختصاصيي المعلومات وجهدهم في وضع اللغات المقيدة، بما ينطوي عليه من تحكم، من شأنه أن يؤدي إلى شيء من الشوشرة في نظام التواصل بين المصدر أو المنشئ من جهة والقارئ أو المتلقي من جهة أخرى، نتيجة لإضافة بعض الملامح أو الخواص الاصطناعية، التي قلما تكون محل اتفاق بين الطرفين. (المترجم)

(**) يعني ذلك أنه تظل هناك دائما فجوة معجمية بين ا اللغة الطبيعية التي تستعمل للتعبير عن الأفكار والمفاهيم من منشئها، واللغة المقيدة. فقد تستعمل مصطلحات جديدة في اللغة الطبيعية لا تستوعبها اللغة المقيدة إلا بعد حين طال أو قصر. (المترجم)

قضية التكلفة:

يستنفد إنشاء وصيانة اللغة المقيدة للتنظيم والاسترجاع، واكتساب القدرة على استعمالها وقتاً لا يستهان به. ومن شأن هذا الوقت اللازم لهذه الأنشطة، أن يترجم في النهاية إلى تكلفة في تنظيم المعلومات واسترجاعها. وفي مقابل ذلك، فإن اللغة الطبيعية هي اللغة التي يستعملها البشر بوجه عام. ولا يتطلب تنظيم المعلومات واسترجاعها باللغة الطبيعية أي تدريب أو صيانة.

قضية التوافق:

يحدث في بعض الأحيان أن يصبح نظام اختزان المعلومات واسترجاعها، بحاجة لتغيير لغته في أثناء تطويره، أو عندما يرغب أحد المستفيدين في إجراء عملية البحث نفسها في عدة مراصد بيانات. وهكذا تبرز قضية التوافق أو التناغم compatibility بالنسبة للغة المقيدة؛ نظرا لأن لكل لغة مقيدة، مقوماتها وخصائصها المتميزة. ويبدو من المستحيل، على سبيل المثال، استخدام تصنيف ديوي العشري، لإنجاز مهام تنظيم واسترجاع، في فهرس متاح للجمهور على الخط المباشر (OPAC) اختار تصنيف مكتبة الكونجرس بوصفه لغته المقيدة. وإذا كانت النظم قائمة على اللغة الطبيعية، فإنه يمكن حينئذ للتحول أو الانتقال من نظام إلى آخر، أن يكون سلسا على الدوام، لأن اللغة الطبيعية ينبغي أن تكون دائماً متوافقة مع نفسها. وتسمى هذه القضية بقضية القابلية للتشغيل التبادلي، باستعارة المصطلحات المستعملة في النظم الإلكترونية (Zeng & Chan, 2004). وعلى النحو نفسه أصبح "نظم معالجة المعرفة knowledge organization systems" مصطلحا جديدا بالنسبة للغات المقيدة.

وقصارى القول، فإن مظاهر قوة اللغة المقيدة في الترادف، والمشترك اللفظي، والنَظْم تعد نقاط ضعف بالنسبة للغة الطبيعية. وكذلك الحال أيضا فإن مواطن الضعف في اللغة المقيدة وهي الدقة، والتحديث، والتكلفة، والتوافق، نقاط القوة بالنسبة للغة الطبيعية. وكما كررت راولي (1994) Rowley مراراً في أحد أعمالها:

> هناك اعتراف عام بأن كلا من اللغة المقيدة واللغة الطبيعية ينبغي استعمالهما مرتبطتين معا، كما أن هناك شيء من الاتفاق حول المزايا النسبية لكل من النظامين. (p. 116)

وبعبارة أخرى، فإن كلا من اللغة الطبيعية واللغة المقيدة، قد وجدت مكانها في تنظيم المعلومات واسترجاعها، ولكن هل سيستمر الأمر كذلك؟

اللغة لأغراض تنظيم المعلومات واسترجاعها في العصر الرقمي:

تستخدم كل من اللغة المقيدة واللغة الطبيعية بالتوازي، لتنظيم المعلومات واسترجاعها في البيئة الورقية. وفي عالم الخط المباشر، لا تزال هناك حاجة للتدخل البشري العميق في تنظيم المعلومات واسترجاعها؛ إذ تحتل اللغة المقيدة الموقع المسيطر كلغة للتنظيم والاسترجاع. وفي العصر الرقمي أصبحت اللغة الطبيعية هي القاعدة بالنسبة لتنظيم المعلومات واسترجاعها، بينما لا تستعمل اللغة المقيدة إلا عرضا أو على نطاق ضيق جدا، نظرا للخواص الفريدة للمعلومات الرقمية.

وتتوافر على الإنترنت نسبة مئوية ضخمة من المعلومات الرقمية، بينما تبدو تقنيات ضبط الجودة غائبة، ومتوسط العمر الافتراضي للمعلومات قصير، في حين ينمو كم المعلومات أسياً(*). ومن ثم فإن أي استخدام للغة المقيدة، باهظة التكلفة في

(*) يتضاعف كل فترة زمنية معينة، أي ينمو بمتوالية هندسية. (المترجم)

البيئة الرقمية لتنظيم المعلومات واسترجاعها، يتطلب تبريرا وترشيدا. ولا تستخدم نظم استرجاع الإنترنت الشهيرة، التي تتكفل بمعظم مهام تنظيم المعلومات واسترجاعها على الإنترنت، اللغة المقيدة على الإطلاق. ويبدو استعمال قوائم الكلمات من جانب قليل من نظم استرجاع الإنترنت، أقرب شيء للغة المقيدة في تنظيم المعلومات الرقمية واسترجاعها.

غير أن اللغة الطبيعية لا يمكن أن تكون اللغة الوحيدة لتنظيم المعلومات واسترجاعها في البيئة الرقمية، نظرا لأن الافتقار إلى ضبط اللغة أو التحكم فيها يمكن أن يكون واحدا من الأسباب الجوهرية لإنخفاض جودة نتائج أو مخرجات نظم استرجاع الإنترنت. ويمكن بوجه عام لمهمة ضبط اللغة أن تتحول من عبء على كاهل اختصاصيي المعلومات، إلى كاهل المستفيدين النهائيين، عندما يقع الاختيار على اللغة الطبيعية لتنظيم المعلومات واسترجاعها. إلا أنه عند استعمال اللغة المقيدة رسميا، فإن مهمة ضبط اللغة ينبغي أن تقع على عاتق اختصاصيي المعلومات. وما لم يحدث ذلك، فإن المستفيد النهائي ينبغي أن ينهض بدور ضبط اللغة على نحو غير رسمي، وذلك على سبيل المثال، بالتفكر في المصطلحات المترادفة التي تدعو الحاجة إليها، في صياغة استفسارات البحث. ولا ينبغي أن يصبح التفاعل المتزايد بين المستفيد النهائي ونظام استرجاع المعلومات، في البيئة الرقمية، ذريعة لإلقاء عبء ضبط اللغة على كاهل المستفيد النهائي؛ نظرا لأنه لا يمكن حتى لاختصاصيي المعلومات، سوى النهوض بمهمة متواضعة في هذا الصدد.

إذن، كيف يمكن أن يكون مستقبل اللغة المقيدة كلغة لتنظيم المعلومات واسترجاعها في العصر الزقمي؟ ولا تبدو الإجابة عن هذا السؤال مؤكدة على الإطلاق. وبنظرة فيما مضى يبدو أن هناك أربع طرق مختلفة يمكن بها استعمال اللغة المقيدة في تنظيم المعلومات واسترجاعها (Lancaster & Warner, 1993):

١. استعمال اللغة المقيدة لكل من التنظيم والاسترجاع.

٢. استعمال اللغة الطبيعية لكل من التنظيم والاسترجاع، ولكن بمساعدة أدوات اللغة المقيدة، كالأدوار والربط المسبق.

٣. استعمال اللغة المقيدة للتنظيم فقط، أما الضبط من أجل الاسترجاع فيتم بوسائل أخرى كاللغة المقيدة الكامنة أو غير المرئية في النظام.

٤. استعمال اللغة المقيدة للاسترجاع فقط. ويتم تطبيق هذا النهج باستعمال مكنز البحث فقط، الذي يعرف أيضا بلغة الربط اللاحق.

وعند النظر في خصائص تنظيم المعلومات واسترجاعها في البيئة الرقمية، يبدو النهج الثاني أكثر جدوى من غيره. فكل من النهج الثالث والرابع يختزنان لغة مقيدة على الخط المباشر من أجل المراجعة أو البحث، الأمر الذي يبدو أيضا خيارا مناسبا لاستخدام اللغة المقيدة. ومن ناحية أخرى، شهد مجال تنظيم المعلومات واسترجاعها في السنوات الأخيرة، بروز أطر لغوية كالتقسيمات الهرمية taxonomies، والتصنيفات المتعارف عليها folksonomies، والتصنيفات التخصصية ontologies. وعلى الرغم من أن لكل إطار خصائصه الفريدة، فإنها جميعا قد تطورت لأغراض التنظيم والاسترجاع في العصر الرقمي.

التقسيمات الهرمية:

نشأ مصطلح التقسيم الهرمي taxonomy، الذي ناقشناه بإيجاز في ٢/ ٢، كتركيب مزجي من الكلمتين اليونانيتين taxis (بمعنى نظام أو ترتيب) وnomos (بمعنى قانون أو علم). وكان التقسيم الهرمي يدل في البداية على تصنيف الكائنات الحية (كالحيوانات والنباتات على سبيل المثال). إلا أن هذا المصطلح يكتسب الآن معنى أكثر اتساعا وأكثر عمومية مما كان من قبل؛ إذ يعرف الآن بأنه تصنيف جميع

الأشياء، كما يمتد معناه ليتجاوز حدود التغطية السابقة للكائنات الحيـة. وكـما ذهـب جلكرست (2003) Gilchrist فإن أقـدم إشـارة إلى الاسـتعمال الحـديث للمـصطلح، ظهرت عام ١٩٩٧ في مقالة حول ياهو! !Yahoo، أحد نظم استرجاع الإنترنت، الذي يستمد جذوره من خدماته الإرشادية القائمة على التقسيم الهرمي الذي اشتهر به.

وتمتد جذور التقسيمات الهرمية في الواقع إلى كل من خطط التصنيف والمكـانز. وكما هو الحال في خطط التصنيف، فإنها تعبر عن العلاقات الهرمية فيما بين الفئات، أي الوحـدات الأسـاس للتقـسيم الهرمـي، بمـستويات مختلفـة للعـرض، ولا تـستعمل الأشكال الهجائية الرقمية للرموز، وسيلة التصنيف المألوفـة، في التقـسيمات الهرميـة. ويتم التعبير عن العلاقات الترابطية فيما بين الفئات هجائيا عند كل مستوى، على نحو يحاكي بنيان إحـدى طـرق العـرض المألوفـة في المكـانز. إلا أنـه عـلى عكـس خطـط التصنيف والمكانز، لا تكفل التقسيمات الهرمية أي وسيلة للإحالة، الأمر الذي ينتقص من الوظيفة التي يفترض للغة المقيدة أن تتكفل بها.

وتستخدم التقسيمات الهرمية التي تتحدد معالمها بخواصها، بوجه عام لتقـسيم المعلومات إلى فئات، ودعم التصفح، أحد طريقتي الاسترجاع. وتبدو هذه التقسيمات جاذبـة بوجـه خـاص بالنـسبة لمرافـئ portals معلومـات المـشروعات، مـن أجـل احتياجاتها لتنظيم المعلومات واسترجاعها (Gilchrist, 2003)، بالإضافة إلى تنفيذها على الإنترنت. ومن بين الأسـباب الكامنـة وراء هـذا الـضرب مـن الاهـتمام، قـدرة التقسيمات الهرمية على استيعاب اللغات المختلفة، التي يمكن أن تتـوافر في المـشروع. يضاف إلى ذلك أن بناء تقسيم هرمي وصيانته عادة ما يكون أيسر نسبيا وأقـل تكلفـة من توفير لغة مقيدة تقليدية كالمكنز مثلا، عندما تكون عمليات الدمج والإضافة، تتم بكثافة في مجال نشاط المشروع. وقد درس كـل مـن وانـج وتـشوذري وخـو Wang, Chaudhry and Khoo (2006) التوقعـات المحتملـة الأخـرى بالنـسبة للتقـسيمات

الهرمية في تنظيم المعلومات واسترجاعها. إلا أن أشهر مثال للتقسيم الهرمي، هو ذلك المستخدم في تجميع دليل ياهو! Yahoo! Directory (dir.yahoo.com).

التصنيفات المتعارف عليها:

كما ألمحنا في ٢/ ١/ ٤، فإن التصنيف المتعارف عليه folksonomy، يمكن النظر إليه بوصفه تقسيما هرميا ينشئه الجمهور أو المستفيدون (Vander Wal, 2007). ومن ناحية أخرى يبدو التصنيف المتعارف عليه متصلا اتصالا وثيقا بالتمييز الاجتماعي social tagging، وهو في جوهره، ناتج جانبي للتمييز الاجتماعي، نظرا لأنه يبنى بتيجان tags يختارها المستفيدون أثناء عملية التمييز.

وعادة ما تتخذ التصنيفات المتعارف عليها شكل سحابة التيجان tag cloud، وهي تجمع مرئي للتيجان التي يتم إنشاؤها، في موقع تمييز قائم على ترددات التيجان. وسحب التيجان، والتصنيفات المتعارف عليها، إذا أردنا الدقة مترادفان. وعلى عكس التقسيمات الهرمية، فإن التصنيفات المتعارف عليها لا تظهر أي علاقات هرمية فيما بين مكوناتها، وهي التيجان على وجه التحديد. وما تتم المحافظة عليه في التصنيفات المتعارف عليها، هو العلاقة الترابطية بين التيجان، مسجلة بترتيب هجائي، دون أي من رموز الإحالات، التي غالبا ما تطبق في اللغة المقيدة التقليدية (كالمكنز مثلا). ومن ثم فإن التصنيفات المتعارف عليها لا يمكن أن تعامل بالطريقة نفسها، التي تعامل بها اللغات المقيدة الأخرى، التي تناولناها فعلا، من حيث تكوينها، وبنيانها ووظيفتها.

وفضلا عن ذلك، فإن جميع اللغات المقيدة، من خطط التصنيف إلى التقسيمات الهرمية، عادة ما يتم تطويرها بواسطة اختصاصيي المعلومات. وفي مقابل ذلك فإن التصنيف المتعارف عليه ضرب جديد من لغات تنظيم المعلومات واسترجاعها، ينشئه المستفيدون النهائيون لأنفسهم دون سواهم، في حركة الجيل الثاني للعنكبوتية Web

2.0، التي لا حدود لها من حيث الموضوع أو الثقافة أو الجغرافيا (Munk & Mørk, 2007). وفي أثناء التمييز يمكن للمستفيدين اختيار تيجان من أي تصنيف متعارف عليه، ولكن تظل لهم حرية اختيار أي مصطلحات من لغتهم هم كتيجان. ونظرا لأن جميع التيجان في التصنيفات المتعارف عليها تربطها روابط فائقة، فإن هذا النظام يكفل للمستفيدين القدرة على تصفح ما هو متاح في موقع تمييز ما، عن طريق تيجان مرتبطة بروابط فائقة، إلى جانب البحث أو التنقيب.

وقد نوقشت مزايا وأوجه قصور التصنيف المتعارف عليه كلغة لتنظيم المعلومات واسترجاعها، باستفاضة من جانب كثير من الباحثين (Noruzi, 2006; Speller, 2007; Spiteri, 2007; Trant, 2006 على سبيل المثال)، وذلك من منظور التصنيفات المتعارف عليها فقط، أو بالإشارة إلى اللغات المقيدة التقليدية. وقصارى القول، فإن التصنيفات المتعارف عليها قد ورثت جميع ما للغة الطبيعية وما عليها، بالإضافة إلى إحدى الخواص الإضافية للغة المقيدة، ألا وهي الترتيب الهجائي visulaization للتيجان، القائم على التردد. وبذلك المعنى، فإن التصنيف المتعارف عليه يستعمل كلغة طبيعية أكثر من استعماله كلغة مقيدة، في البيئة الرقمية لتنظيم المعلومات واسترجاعها. ومن الممكن العثور على مثال للتصنيف المتعارف عليه في موقع Delicious.com عندما يتم تصوير إحدى سحب التيجان.

التصنيفات التخصصية:

يهتم مصطلح علم الوجود ontology، المستعار من مجال الفلسفة، بدراسة الكينونة أو الوجود. وقد لجأ الباحثون في الذكاء الاصطناعي إلى هذا المصطلح في ثمانينيات القرن العشرين، للدلالة على اكتساب المعرفة والتعبير عنها، عندما يكون المرء بصدد وضع تصور نظري لمجال ما، في تصميم وتنفيذ النظم الخبيرة (Vickery, 1997). وعادة ما يعرف الوجود أو التصنيف التخصصي، في هندسة المعرفة، أو في

علم الحاسبات والمعلومات بوجه عام، بالتحديد الصريح الرسمي لمعالم تصور مفاهيمي أو نظري conceptualization مشترك (Gruber, 1993).

ومن ناحية أخرى تبنى برنرز-لي، وهندلر، ولاسيلا Berners-Lee, Hendler, and Lassila (2001) هذا المصطلح، عند وضع تصور العنكبوتية الدلالية. وهذا موضوع نستكشفه في ١٢/ ٣. وفي هذه البيئة على وجه التحديد، ينفصل الوجود فعلا عن أصله في الفلسفة، ويحظى بالترحيب بوصفه لغة لتنظيم المعلومات واسترجاعها. وفي بحثهم البذري، يتناول برنرز-لي، وهندلر، ولاسيلا Berners-Lee, Hendler, and Lassila (2001) خصائص الخطط التخصصية ontologies بوصفها:

> مجموعات من العبارات المصاغة بلغة كإطار وصف المصادر RDF على سبيل المثال، تحدد معالم العلاقات بين المفاهيم، وتحدد القواعد المنطقية لتبريرها؛ إذ يصبح بإمكان الحاسبات "فهم" معنى البيانات الدلالية على صفحة العنكبوتية؛ بتتبع الروابط التي تفضي إلى معلومات بعينها. (P. 38)

ويمكن للخطط التخصصية أن تتخذ أشكالا مختلفة. وقد أحصى أوشولد Uschold (1996) الأشكال الأربعة التالية، التي تتراوح ما بين المتحللة من الرسميات إلى حد بعيد، والرسمية على نحو صارم، وذلك من منظور هندسة المعرفة. ويتم التعبير عن الخطط المتحللة من الرسميات إلى حد بعيد بطريقة غير منضبطة loosely باللغة الطبيعية. ويأتي بعد ذلك على الطيف، الخطط التخصصية غير الرسمية المنتظمة، التي تستعمل لغة طبيعية محدودة ومنتظمة، لزيادة الوضوح والحد من الغموض. ويسمى الشكل الثالث المعارف شبه الرسمية semiformal ontology، ويتم التعبير عنها بلغة اصطناعية محددة شكلا. أما الشكل الأخير في هذا

الطيف فهو المعارف الرسمية على نحو صارم، التي يمكن أن تحدد بشكل شديد الدقة في التفاصيل، معالم المصطلحات بعلم الدلالة الرسمي والنظريات الجزئية theorems. وعلى الرغم من غياب الوصف المعترف به لشكل التصنيف التخصصي على ضوء العنكبوتية الدلالية، فإن الشكل المرشح ربما يكون النوع الرسمي على نحو صارم الذي رسم معالمه أوشولد (1996) Oschold.

وتشمل العلاقات بين المفاهيم في الخطط التخصصية، الترادف والتضاد، والكلمات أو العبارات التي يدخل مجالها الدلالي ضمن المجال الدلالي لأخرى hyponomy (علاقة الكينونة "is-a") والتقسيم الهرمي الذي يعالج علاقة الجزء بالكل meronomy. وغالبا ما يتم التعبير عن تلك العلاقات بالأشكال البيانية لإطار وصف المصادر RDF، لإنشاء العنكبوتية الدلالية (Gilchrist, 2003). يضاف إلى ذلك أن الخطط التخصصية ينبغي أن تحدد القواعد المنطقية للتبرير، حول المفاهيم والعلاقات الداخلة فيما بينها، نظرا لأنها على عكس اللغات المقيدة التقليدية (كالمكانز مثلا) ليست ثابتة؛ إذ ينبغي أن يتوافر لها التقنية التي تدل على التغيرات، وتنجز التجديدات آليا. ويمكن للخطط التخصصية بالارتباط بالمكونات الأخرى للعنكبوتية الدلالية، أن تجعل الحاسبات تستوعب الدلالات التي تنطوي عليها مصادر العنكبوتية.

وفضلا عن ذلك، فإن وظيفية الخطط التخصصية، من الواضح أنها ليست هي نفسها وظيفية اللغات المقيدة التقليدية (كالمكانز وخطط التصنيف مثلا)؛ فالأولى يقصد بها في المقام الأول، نقل الدلالات من مصادر العنكبوتية إلى الحاسبات، لا ضبط استعمال المصطلحات في تنظيم المعلومات واسترجاعها. ونورد في ١٢/ ٣ المزيد من مناقشات الخطط التخصصية، ذلك القسم المكرس للعنكبوتية الدلالية. وغالبا ما ينظر إلى ورد نت Word Net (wordnet.princeton.edu) بوصفها مثالا للخطط التخصصية.

واللغة اللازمة لتنظيم المعلومات واسترجاعها في العصر الرقمي، ناشئة بالتأكيد، ويتم إجراء البحوث بكثافة لمعالجة قضية اللغة في المجال. وتحويل اللغات vocabulary switching، أحد تقنيات التحول آليا من إحدى لغات تنظيم المعلومات واسترجاعها إلى أخرى، عبر مختلف المجالات الموضوعية، من بين مجالات البحث والاستكشاف، كخطة واعدة لحسم الجدل حول اللغة الطبيعية في مقابل اللغة المقيدة. وبمجرد أن يتم تنفيذ هذه الخطة، سوف يقع الاختيار دون تردد على استعمال لغة لتنظيم المعلومات واسترجاعها، دون الاقتصار على اختيار محدد سلفا، ولن يقتصر الأمر على اللغة في مجال موضوعي بعينه، إذ يمكن للمستفيدين كسر استرقاق أو استعباد تخصصاتهم القريبة، والإفادة من المعلومات العلمية بأسرها فعلا، في البحث (Schatz, 1997). ويختلف تحويل اللغات عن "اللغة المقيدة غير المرئية" من ناحيتين؛ أولاهما أن تحويل اللغات يعتمد بكثافة على البحث في معالجة اللغة الطبيعية، والأخرى أن تحويل اللغات يتعامل مع لغات تنظيم المعلومات واسترجاعها في مجالات عدة (أكثر من اثنين بكثير) بينما اللغة المقيدة غير المرئية تتعامل في الأساس مع الترجمة بين اللغة الطبيعية وإحدى اللغات المقيدة المختزنة على الخط المباشر. فقد أنشأ تشن وشاتز Chen, Schatz وزملاؤهما، على سبيل المثال، فضاءات مفاهيم concept spaces لعشرة ملايين من مستخلصات المقالات، موزعة على ألف مجال موضوعي، تغطي العلوم والتقنية بأسرها (Schatz, 1997). وفضلا عن ذلك تبين أن فضاءات المفاهيم فعالة بالنسبة للاقتراح التفاعلي للمصطلحات، وتحويل اللغات (Chen, et al., 1997).

وقصارى القول، فإن التنظيم والاسترجاع الآليين، بدلالات قابلة للامتداد والتوسع scalable، يبدوان هما المستقبل، مادامت قضية اللغة في العصر الرقمي، هي الشغل الشاغل. ويمكن لتتابع الأحداث أن يصبح حقيقة إذا ما تطورت العنكبوتية الدلالية التي تصورها برنرز-لي Berners-Lee وزملاؤه.

المراجع

Aitchison, Jean, Gilchrist, Alan, and Bawden, David. (1997). *Thesaurus construction and use: A practical manual.* 3rd ed. London: Aslib.

Burners-Lee, Tim, Hendler, James, and Lassila, Ora. (2001). The Semantic web. *Scientific American,* 284(5), 35-43.

Chen, Hsinchun, et al. (1997). A concept space approach to addressing the vocabulary problem in scientific information retrieval: An experiment on the Worm Community System. *Journal of the American Society for Information Science,* 48(1), 17-31.

Gilchrist, Alan. (2003). Thesauri, taxonomies and ontologies: An etymological note. *Journal of Documentation,* 59(1), 7-18.

Gruber, Tom R. (1993). A translation approach to portable ontology specifications. *Knowledge Engineering,* 5(2), 199-220. retrieved November 20, 2008, from www.ksl.stanford.edu/kst/what-is-an-ontology.html

Lancaster, F.W. (1986). *Vocabulary control for information retrieval.* 2nd ed. Arlington, VA: Information Resources Press.

Lancaster, F. W., and Warner, Amy J. (1993). *Information retrieval today.* Arlington, VA: Information Resources Press.

Milstead, Jessica L. (1995). Invisible thesauri: The year 2000. *Online & CD ROM Review,* 19(2), 93-94.

Munk, Timme Bisgaard, and Mørk, Kristian. (2007). Folksonomy: the power law and the significance of the least effort. *Knowledge Organization,* 34(1), 16-33.

National Information Standards Organization. (1993). *Guidelines for the construction, format, and management of monolingual thesauri* (ANSI/NISOZ39.19-1993). Bethesda, MD: NISO Press.

Noruzi, Alireza (2006). Folksonomies: (Un) Controlled vocabulary? *Knowledge Organization,* 33(4), 199-203.

Rowley, Jennifer E. (1992). *Organizing knowledge: An introduction to information retrieval.* Brookfield, VT: Gower.

Rowley, Jennifer E. (1994). The controlled versus natural indexing languages debate revisited: A perspective on information retrieval practice and research. *Journal of Information Science,* 20(2), 108-119.

Schatz, Bruce R. (January 17, 1997). Information retrieval in digital libraries: Bringing research to the net. *Science,* 275, 327-334.

Speller, Edith. (2007). Collaborative tagging folksonomies, distributed classification or ethnoclassification: A literature review. *Library Student Journal,* (2)1. Retrieved November 15, 2008, from www.librarystudentjournal.org/index.php/ls/article/view/45/59.

Spiteri, Louise F. (2007). The structure and form of folksonomy tags: The road to the public library catalog. *Information Technology & Libraries,* 26(3), 13-25.

Trant, J. (2006). Exploring the potential for social tagging and folksonomy in art museums: Proof of concept. *New review of Hypermedia and Multimedia,* 12(1), 83-105.

Uschold, Mike. (1996). Building ontologies: Towards a unified methodology. *Proceedings of the 16th Annual Conference of the British Computer Society Specialist Group on Expert Systems.* Cambridge, England, December 16-18, 1996. Retrieved November 24, 2008, from www.aiai.ed.ac.uk/project/ftp/documents/1996/96-es96-unifiled-mehtod.pdf

Vander Wal, Thomas. (2007). Folksonomy. Retrieved November 20, 2008, from vanderwal.net/folksonomy.html

Vickery, B. C. (1997). Ontologies. *Journal of Information Science,* 23(4), 277-286.

Wang, Zhonghong, Chaudhry, Abdus Stattar, and Khoo, Christopher. (2006). Potential and prospects of taxonomies for content organization. *Knowledge Organization,* 33(3), 160-169.

Wellisch, Hans H. (1995). *Indexing from A to Z.* 2nd ed. New York: H.W. Wilson.

Zeng, Marcia L., and Chan, Lois Mai. (2004). Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology,* 55(5), 377-395.

الفصل الخامس

تقنيات الاسترجاع وصياغة الاستفسارات

الآن، وقد ناقشنا قضية اللغة في تنظيم المعلومات واسترجاعها في الفصل الرابع، آن الأوان للانتقال إلى الموضوع التالي، وهو تقنيات الاسترجاع التي ينبغي وضعها في الحسبان، عند إجراء عمليات البحث. ونتناول بعد ذلك بالشرح، التعبير عن الاستفسارات أو صياغة الاستفسارات.

تقنيات الاسترجاع:

يقصد بتقنيات الاسترجاع مساعدة المستفيدين في العثور على ما يحتاجون إليه من معلومات، بفاعلية وكفاءة. وتتوافر للمستفيدين تقنيات متعددة، نتيجة لتطور تقنيات المعلومات، وتقدم البحث العلمي في تقنيات الاسترجاع. وهناك مجموعة من تقنيات الاسترجاع تسمى بالطرق الأساس، بينما تسمى التقنيات الأخرى الطرق المتقدمة.

تقنيات الاسترجاع الأساس:

تقنيات الاسترجاع، كالبحث البوليني **Boolean searching**، ومراعاة الحرف الأول من الكلمة **Case sensitivity**، والبتر **Truncation**، والبحث بالتقارب **Proximity searching**، والبحث بالحقل **Field search**، تقنيات أساس؛ نظرا لأنها تحظى بوجه عام بالدعم من جانب معظم نظم استرجاع المعلومات. ونتناول في الأقسام التالية وظيفة ومقومات هذه التقنيات الأساس كلا على حدة.

البحث البوليني:

كما أشرنا بإيجاز في ١/ ١/ ٢/ ١، سمي البحث البوليني باسم عالم الرياضيات الإنجليزي جورج بول **George Boole**، الذي أجرى تحليلات رياضية للمنطق. إذ استعمل بول ثلاثة معاملات، هي "و **AND**" و"أو **Or**"، و"فيما عدا **NOT**" لتلخيص العمليات المنطقية للعقل البشري، وهي جمع (و**AND**) مفاهيم مختلفة لتكوين مفاهيم أكثر تعقداً، وفصل (فيما عدا **NOT**) المفاهيم المركبة في مفاهيم مفردة أكثر بساطة (**Smith, 1993**)، وتضمين (أو **OR**) المزيد من المفاهيم لتوسعة دلالتها. وتعرف كل هذه العمليات المنطقية أيضاً بالجبر البوليني. وعندما يستخدم الجبر البوليني في استرجاع المعلومات، تنهض المعاملات الثلاثة المسماة بالمعاملات البولينية، بالوظائف التالية:

- المعامل و"**AND**" للحد من اتساع عملية البحث.
- المعامل أو" **O**" لتوسعة عملية البحث.
- المعامل فيما عدا "**NOT**" لاستبعاد النتائج غير المرغوب فيها.

ويجمع المعامل "و **AND**" بين مصطلحين أو أكثر، في استراتيجية البحث، ويشترط وجود جميع المصطلحات في الوثائق التي يتم العثور عليها. ففي استراتيجية البحث، التصفية والجدال **Filtering AND Controversy** على سبيل المثال، ينبغي الحصول فقط على النتائج حول التصفية بوصفها جدالاً. ولا يمكن لأي نتائج حول التصفية بوجه عام، أو أي مجادلات أخرى، كالتحكم في السلاح، أن تلبي شروط البحث. ويفيد المعامل و"**AND**" على وجه الخصوص في البحث عن معلومات حول مفاهيم مركبة.

أما المعامل "أو OR" فيوسع عملية البحث بإدخال المزيد من المصطلحات في استراتيجية البحث. وغالباً ما تكون المصطلحات التي يتم الربط بينها معاً بـأو OR مترادفات أو مصطلحات متصلة. وينظر إلى الوثائق التي يرد بها أي من هذه المصطلحات بوصفها تسديدات Hits ملائمة أو نتائج متوقعة. فمن الممكن على سبيل المثال لاستراتيجية بحث مكونة من "التصفية أو الجدال" أن تسفر عن نتائج حول جميع أنواع التصفية وجميع أنواع المجادلات، بالإضافة إلى الوثائق حول التصفية بوصفها جدالاً. ومن الواضح أن المعامل "أو OR" يهدف إلى توسعة مجال عملية البحث، ويزيد من ناتج الاسترجاع.

أما المعامل "فيما عدا NOT". فأكثر تقييدا بالمقارنة بكل من "و" و "أو"؛"ففيما عدا" يقيد مجال عملية البحث، باستبعاد المصطلح أو المصطلحات التي ترد بعد المعامل "فيما عدا"؛ فمن الممكن، على سبيل المثال، لاستفسار البحث "التصفية فيما عدا الجدال" أن يسترجع معلومات حول جميع أنواع التصفية (كتنقية الماء وتنقية الهواء على سبيل المثال) مادامت ليست جدالاً. ومن الممكن استعمال معامل "فيما عدا NOT" عمداً لاستبعاد النتائج غير المرغوب فيها، عند صياغة استفسار البحث.

وعندما يتم استعمال نوع واحد فقط من المعاملات البولينية في استراتيجية بحث ما، تسمى العملية بالبحث البوليني البسيط. وعندما تشتمل استراتيجية البحث الواحدة على أكثر من نوع واحد من المعاملات البولينية، تسمى العملية بالبحث البوليني المركب. والترتيب الطبيعي بالنسبة لمعالجة المعاملات البولينية، في معظم نظم استرجاع المعلومات هو:

- أولاً، معامل فيما عدا NOT.
- ثانياً، معامل وAND.
- ثالثاً، معامل أو OR.

ففي استراتيجية البحث البوليني المركب "التصفية أو الرقابة والجدال فيما عدا المكتبات" على سبيل المثال، يمكن لمعامل فيما عدا (أي الجدال فيما عدا المكتبات) أن يعالج أولاً، لاستبعاد أي وثائق حول الجدال المتعلق بالمكتبات، ثم يجمع المعامل و**AND** المصطلحين "الرقابة" و"الجدال" مع مفهوم "المكتبة" الذي استبعد فعلاً. وأخيراً يتم ربط مصطلح "التصفية" بعلاقة أو**"OR"** بمجموعة الوثائق التي أسفرت عنها عمليتا البحث السابقتان. ومالم تكن نتائج عملية البحث تلبي توقعات المستفيد النهائي، فإنه يمكن استعمال الأقواس لتغيير الترتيب الطبيعي للمعالجة، أو تحديد الترتيب المقصود للمعالجة. ففي استراتيجية البحث السابقة، على سبيل المثال، يمكن للمستفيد النهائي أن يستخدم الأقواس على هذا النحو: (التصفية أو الرقابة) و(الجدال) فيما عدا المكتبات. ومن ثم فإن ترتيب المعالجة بالنسبة لهذه الاستراتيجية، يمكن أن يكون أو**"OR"**، وو**""AND** ثم فيما عدا **"NOT"**، لأن المعالجة يمكن أن تتبع القاعدة التي تقضي بأن تتم معالجة الأقواس الداخلية أكثر من غيرها **innermost** أولاً، تليها الأقواس الخارجية، وهكذا. ومن شأن الوثائق التي يتم استرجاعها أن تكون حول "التصفية" أو "الرقابة" بوصفها "جدالاً"، ولكنها لا تتصل "بالمكتبات". وكما هو مبين، فإنه يمكن استعمال ثنائيات متعددة من الأقواس لتحديد ترتيب بعينه للمعالجة، في استراتيجية البحث البوليني المركب. ولهذا السبب أحياناً ما يسمى البحث البوليني المركب بالبوليني المسيج **nested**.

وكأسلوب للاسترجاع، فإن البحث البوليني هو الأسلوب الذي يستعمل إلى حد بعيد، في الغالب الأعم، من جانب جميع نظم الاسترجاع القائمة. إلا أن الأمر يتطلب قدراً من الوقت والمهارسة للتمكن من أساليب البحث البوليني، وخصوصاً أسلوب البحث البوليني المركب.

البحث بمراعاة الحرف الأول من الكلمة:

بالنسبة للغات كالإنجليزية والفرنسية والإسبانية، التي تحدث فيها الأحرف الكبيرة والأحرف الصغيرة في بداية الكلمات نوعاً من الاختلاف، يكفل البحث بمراعاة الحرف الأول من الكلمة **Case-sensitive** للمستفيد النهائي القدرة على أن يعين على وجه التحديد، كيف يتم تمثيل المصطلح في الاستفسار والنظام؛ فالكلمتان **Target** و**target**، على سبيل المثال، تدلان على أمرين مختلفين. فالأولى التي تبدأ بحرف كبير، تدل على متجر للبيع بالتجزئة في الولايات المتحدة، بينما الثانية التي تبدأ بحرف صغير، ببساطة مصطلح عام يعني الهدف أو الغاية وما شابه ذلك. وبمجرد أن يتم التمييز بين الشكلين، يمكن للمستفيد أن يقرر ما إذا كانت هناك حاجة للبحث بمراعاة الحرف الأول. فإذا كان المرء ينوي البحث عن معلومات حول متجر تارجت، فإن كلمة **Target** التي تبدأ بالحرف الكبير ينبغي أن يقع عليها الاختيار كمصطلح للبحث. ومالم يكن الأمر كذلك فإنه يمكن استعمال الكلمة **target** التي تبدأ بحرف صغير.

ويطبق البحث بمراعاة الحرف الأول من الكلمة على نطاق ضيق جداً، بالمقارنة بالبحث البوليني. إلا أن البحث بمراعاة الحرف الأول من الكلمة ينجز أنواعاً معينة من مهام الاسترجاع، لا تستطيع تقنيات الاسترجاع الأخرى إنجازها.

البتـر:

البتر الذي يعرف أحيانا بالبدل وتغيير الشكل **wildcard**، والرد إلى الجذع **stemming**، والتجريد **stripping**، وإخفاء معالم المصطلحات **term masking**، وخوارزمية الخلط **conflation algorithm**، تقنية لاسترجاع الأشكال المختلفة للمصطلحات، التي يجمعها معاً جزء معين. وعادة ما تحدد نظم استرجاع المعلومات رمزاً بعينه (كعلامة الاستفهام؟ أو النجمة * على سبيل المثال) لأغراض الدلالة على

البتر. فاستعمال المصطلح المبتور (*) **network**، على سبيل المثال، كاستفسار يسترجع الوثائق حول **networks** والوثائق حول **networking**.

ويسمى بتر عجز أو كاسعة **suffix** المصطلح، كما سبق أن أشرنا، ببتر الجانب الأيمن، الذي يتكفل بمعظم عمليات البتر. كذلك يمكن للبتر أن يتم أيضاً باستبعاد صدر الكلمة **prefix** أو حشوها **infix** (أي بعض الأحرف التي ترد في منتصف المصطلح). ويسمى حذف الصدر ببتر الجانب الأيسر، بينما يسمى حذف حشو المصطلح بالبتر الأوسط أو الداخلي. وقلما نجد بتر الجانب الأيسر، كما أنه لا يدعمه الآن سوى قليل من نظم استرجاع المعلومات. ومن الممكن استعمال البتر الأوسط لمعالجة المصطلحات التي يحتمل أن تكتب بأكثر من هجاء واحد. فمن الممكن، على سبيل المثال لمصطلح الاستفسار المبتور **Col*r** أن يسترجع كلا من **color** و**colour**. وكذلك الحال أيضاً يمكن للبحث بناء على **organi*aton** أن يسترجع معلومات عن كل من **organisation, organization**.

ومن المهم بتر القطاع الصحيح من المصطلح، إذا ما كان المرء ينوي استعمال أسلوب الاسترجاع هذا. ومن الممكن أن يؤدي البتر المبالغ فيه (كبتر **cotalog** إلى **cat*** على سبيل المثال) أن يؤدي إلى استرجاع قدر كبير من المعلومات غير المرغوب فيها، ومن ناحية أخرى يمكن للبتر المحدود جداً (كبتر **catalogue** إلى **catalogu** على سبيل المثال) ألا يحقق الهدف من البتر. ولتوفير قدر من التحكم في البتر، يكفل كثير من نظم استرجاع المعلومات للمستفيد أيضاً، تحديد العدد الدقيق للأحرف التي يتم بترها(*).

(*) يتصل كل ما ورد عن البتر هنا باللغات التي تكتب من اليسار إلى اليمين. ويتصل البتر ببنية الكلمة أو الصرف Morphology. وتختلف قواعد الصرف من لغة إلى أخرى؛ إذ تختلف اللغات الهندية الأوروبية الغروية أو اللصقية كالإنجليزية مثلاً، عن اللغات السامية الاشتقاقية ومن بينها العربية. (المترجم)

البحث بالتقارب:

بينما يمكن للمعامل وAND أن يبين أي المصطلحات ينبغي أن ترد في نتائج البحث، فإنه لا يمكن أن يحدد مدى تباعد مواقع مصطلحات البحث هذه في الوثيقة. ففي صيغة البحث البولينية **filtering AND controversy** على سبيل المثال، يمكن للمصطلحين **filtering** و **controversy** أن يردا متجاورين، أو يفصل بينهما عدة مئات من الكلمات، أو في أي مكان من الوثيقة المستهدفة. ويكفل البحث بالتقارب **proximity searching**، الذي يعرف أيضاً بالبحث بالتجاور **adjacency searching**، للمستفيد القدرة على أن يحدد بدقة مدى التقارب، أو المسافة الفاصلة بين مصطلحي البحث، ومواقعهما النسبية، باستعمال معاملي (مع **WITH**) (وقريب **NEAR**). ويمكن لرموز البحث بالتقارب أن تختلف من نظام إلى آخر.

ويعني معامل (مع **WITH**) أن المصطلحين ينبغي أن يردا في النظام متجاورين، وبالترتيب نفسه الذي يرد في استفسار البحث. فصيغة البحث "المعلومات مع التقنيات" **information WITH technology** على سبيل المثال، تدل على أن الوثائق التي يسترجعها الاستفسار ينبغي أن تشتمل على "تقنيات المعلومات" **information technology** كعبارة(*)، وليس أي شيء آخر (كالمعلومات والتقنيات **information and technology**، أو "معلومات التقنيات" **technology information** على سبيل المثال). يضاف إلى ذلك، أنه يمكن لتوسعة معامل **WITH**، **nWITH** أن تشترط أن يكون المصطلحان المرتبطان بالمعامل في حدود س **n** (**n= 1, 2...n**) من الكلمات فيما بينهما، وبالترتيب المحدد. فمن الممكن على سبيل المثال لصيغة البحث **information 2 WITH technology** أن تسترجع وثائق

(*) مضاف ومضاف إليه هنا، ويمكن للعبارة أن تكون موصوفا وصفة. (المترجم)

حول "المعلومـات والتقنيـات" **information and technology**، و"المعلومـات وتقنيات الشبكات"**information and network technology**، وما شابه ذلك.

ويعمل المعامل (قريب **NEAR**) بطريقـة مماثلـة، أي أن المـصطلحين اللـذين يربط بينهما ينبغي أن يكونا متجاورين. إلا أنه على عكس المعامل (مع **WITH**)، فـإن المصطلحين المرتبطين بالمعامل قريب يمكن أن يردا بـأي ترتيـب مادامـا متجـاورين. فـصيغة البحـث "المعلومـات قريـب مـن التقنيـات" **information NEAR technology**، على سبيل المثال، يمكن أن تسفر عن نتائج حول "تقنيات المعلوماث" **information technology** أو "معلومـــات التقنيـــات" **technology information**. ويشترط المعامل **nNEAR** أن يكون المصطلحان المرتبطان بالمعامـل في حدود س **n** "**n=1,2,....**) من الكلمات من بعضهما البعض، ولكـن بـأي ترتيـب كان. ومن ثم فإن الوثائق التي تشتمل على عينة العبارات التالية، جميعها نتائج ملائمة لصيغة البحث **information 2 NEAR technology**:

المعلومات والتقنيات **information and technology**.

المعلومات وتقنيات الشبكات **information and network technology**.

التقنيات والمعلومات **technology and information**.

التقنيات والمعلومات الإدارية **technology and business information**.

والبحث بالعبارات **phrase searching**، في مقابل البحث بالكلمات **word searching**، ليس سوى أحـد تطبيقـات البحـث بالتقـارب، إذا كانـت المعلومـات الأصلية التي تم تجميعها لنظم استرجاع المعلومات، قد تم تمثيلها بالكلمات فقـط (أي تكشيف الكلمات). ولكي نكون أكثر تحديداً، فإن المعامـل **WITH** يمكـن أن ينجـز البحث بالعبارات على وجه التحديد، بينما تدعم بقيـة معـاملات التقـارب عمليـات

البحث اعتماداً على تجاورات **juxtapositions** مختلفة للمصطلحات الواردة في صيغة البحث. كذلك يمكن للتجاورات أن تمتد خارج مدى س من الكلمات، ولكن في حدود الحقل نفسه أو الفقرة نفسها، في بعض نظم استرجاع المعلومات. فنظام ديالوج **DIALOG**، على سبيل المثال، وهو نظام لاسترجاع المعلومات على الخط المباشر، يكفل للمستفيد القدرة على أن يشترط ضرورة ورود مصطلحات البحث في حدود الحقل نفسه (كحقل العنوان أو حقل المستخلص على سبيل المثال) أو الفقرة نفسها.

البحث بالحقل:

عادة ما يتم تنظيم معلومات الوثائق بخواص مميزة كالمؤلف، والعنوان وتاريخ النشر ونوع الوثيقة. وتسمى هذه الخواص المميزة بالحقول، بمجرد أن يتم التعبير عنها في نظم استرجاع المعلومات(*). وعندما يتم تنظيم المعلومات أو الوثائق بالحقل، فإنه يمكن استرجاعها فيما بعد بالطريقة نفسها. ويقصر البحث بالحقل، بطبيعته عمليات البحث على حقل واحد أو عدة حقول، بدلا من الوثيقة بأكملها.

وينهض البحث بالحقل بوظيفتين؛ أولاهما تحديد نقطة البدء بالنسبة للمعلومات التي يبحث المرء عنها، إذا كانت الخاصة المميزة للحقل (كاسم المؤلف مثلاً) معروفة؛ فإذا كنا نرغب على سبيل المثال في الحصول على معلومات حول أعمال هانز بيتر لون **Hans Peter Luhn**، المتصلة باسترجاع المعلومات، فإننا نبحث بالمصطلح "استرجاع المعلومات **information retrieval**. إلا أننا سرعان ما يتبين لنا أن مثل ذلك البحث يمكن أن يسترجع، ولا شك معلومات أكثر بكثير مما كان متوقعاً، نظراً لأن جميع الأعمال حول استرجاع المعلومات، بما في ذلك أعمال لون،

(*) عادة ما يتم في نظم استرجاع المعلومات اعداد تسجيلة لكل وثيقة يشملها النظام. ويقسم محتوى التسجيلة إلى حقول، يضم كل منها عنصراً من عناصر البيانات المميزة. (المترجم)

سوف يتم استرجاعها كنتائج لعملية البحث. وإذا ما أجرينا عملية بحث باسم المؤلف، أي البحث بحقل المؤلف، فإننا يمكن أن نكون قادرين على استرجاع كل ما ألف (لون) من وثائق حول استرجاع المعلومات فضلاً عن الموضوعات الأخرى. وحينئذ يمكننا استعمال المعامل و**AND** للجمع بين نتائج عملية البحث السابقة ومصطلح "استرجاع المعلومات" لاستبعاد أعمال (لون) في المجالات الأخرى خلاف استرجاع المعلومات.

وبالإضافة إلى ما أوضحنا تواً، يمكن للبحث بالحقل أن يضيق أيضاً مجال عملية البحث على نحو فعال؛ لنفترض أن عملية بحث عن استرجاع المعلومات قد أجريت في نظام مصمم لعلم المعلومات، فإن الناتج يمكن أن يشتمل على عدة آلاف من التسجيلات. وقليل من المستفيدين من يمكن أن تكون لديه الرغبة في الاطلاع على نتائج عمليات بحث بهذه الضخامة. وعند هذه النقطة يمكن لحقول الخواص المميزة كتاريخ النشر مثلاً، أو اللغة، أو نوع الوثيقة، أن تساعد على نحو فعال في تضييق مجال عملية البحث. والبحث بالموضوع أو المفهوم (أي البحث "بحيثيات أو موضوع **aboutnes**" المعلومات، متبوعاً بالبحث بالحقل (أي البحث "بخواص **ofness** " المعلومات)، من الطرق المألوفة لتنقية نتائج الاسترجاع.

ومعظم المعلومات المتوافرة على الإنترنت، لا يتم تنظيمها بحقول الخواص المميزة في نظم استرجاع الإنترنت، نظراً للأسباب التي سبق أن تناولناها في الفصول السابقة. ولهذا فإن البحث بالحقل غير قابل للتطبيق بوجه عام على استرجاع المعلومات من الإنترنت. إلا أن تقنيات الاسترجاع الرئيسة الأخرى التي ناقشناها في هذا القسم، تحظى بالدعم الكلي أو الجزئي من جانب نظم استرجاع الإنترنت، وهذه قضية نستكشفها بمزيد من التفصيل في الفصل الثامن، "نظم استرجاع المعلومات".

أساليب الاسترجاع المتطورة:

يتم تطبيق أساليب الاسترجاع الرئيسة التي سبق تناولها في معظم- إن لم يكن في كل- نظم استرجاع المعلومات. وفي مقابل ذلك تتوافر أساليب الاسترجاع المتقدمة أو المتطورة **advanced**، انتقائياً، في قطاع صغير من نظم استرجاع المعلومات القائمة، وذلك بعد اختبارها في مختبرات البحث العلمي.

البحث المشوش:

يبدو البحث المشوش **fuzzy** متصلاً بالبتر، ولكن مع اختلاف جوهري؛ فبينما يقصد بالبتر استرجاع مختلف أشكال المصطلح، عندما تتقاسم فيما بينها بعض الأجزاء المشتركة، فإن البحث المشوش يقصد به العثور على المصطلحات التي حدثت أخطاء في هجائها عند إدخال البيانات أو إدخال الاستفسار. فمن الممكن لمصطلح **computer** على سبيل المثال أن يحدث خطأ في هجائه ليصبح **compter** أو **comptuer**، أو **compyter**. ومن الممكن للتعرف على الأحرف ضوئياً **optical character recognition (OCR)**، والنصوص المضغوطة، أن يسفرا أيضاً عن محتوى مشحون بالأخطاء. وعلى ذلك فإن البحث المشوش مصمم على وجه الخصوص، لاكتشاف وتصحيح أخطاء الهجاء أو الأخطاء التي تنشأ عن التعرف على الأحرف ضوئياً، أو ضغط النصوص **(Grassman & Frieder, 1998)**.

وس - جرام **N-gram** أسلوب خاص لتطبيق تقنية البحث المشوش. و س - جرام سلاسل متتابعة ثابتة الطول من س (س = ٢، ٣،....س) **(n=2, 3, ...n)** من أحرف المصطلح **(Kowalski, 1997)**، أو مصطلح مفكك إلى أجزاء من الكلمات بحجم س. وإذا ما اتخذنا مصطلح **fuzzy searching** كمثال، فإنه يمكن أن يصبح لدينا:

ثنائي الجرام

- **Bi - grams (n= 2): fu uz zz zy se ea ar ar rc ch hi in ng**

ثلاثي الجرام

- **Tri-grams (n=3): fuz uzz zzy sea ear arc rch chi hin ing**

وهناك أيضاً أنواع أخرى من أعداد الجرامات **n=grams** (مثل الجرامات الرباعية - **quart grams**، أو الجرامات الخماسية **penta-grams**. وتصبح هذه الأعداد من الجرامات س - جرامات، التي ليس من الضروري أن تمثل المعنى الدلالي للمصطلح، هي الوحدات الأساس للتحقق من الأخطاء المحتملة. وحينئذ تستخدم خوارزميات المضاهاة لتحديد ما إذا كانت هناك مضاهاة بين التمثيل والاستفسار **(Grossman & Frieder, 1998)**). وإذا كانت جميع أعداد الجرامات س- جرام بالنسبة لمصطلح ما تضاهي الاستفسار، بينما يكون هناك واحد أو اثنان لا يضاهيان، يتم اكتشاف الخطأ، ويصبح بالإمكان تقديم الناتج الصحيح. وهناك أسلوب آخر مألوف لتطبيق البحث المشوش، وهو المقارنة المعجمية. وذلك بمقارنة كل كلمة في الاستفسار يتم إدخالها في نظام استرجاع المعلومات، مقابل معجم، ويتم حينئذ تصويب أي خطأ يتم اكتشافه.

وبعبارة أخرى، فإن البحث المشوش يمكن أن يتغاضى عن، أو يتسامح مع الأخطاء التي تقع أثناء إدخال البيانات أو إدخال الاستفسارات. ولا يمكن للوثائق التي تشتمل على أخطاء ناتجة عن الهجاء غير الصائب، أو أخطاء التعرف على الأحرف ضوئياً، وما شابه ذلك من أسباب، أن تسترجع بدون تقنية البحث المشوش.

البحث الموزون:

يحدث في بعض الأحيان، أن يرغب المستفيد في التركيز على كل مصطلح في الاستفسار على نحو متفاوت؛ ففي استفسار البحث "التصفية والجدال" **filtering AND controversy**، على سبيل المثال، يمكن أن يكون المستفيد مهتماً أكثر بالجانب الخاص "بالجدال" لا بالجانب الخاص "بالتصفية" في حد ذاتها. وقد صمم البحث الموزون **weighted** لهذا الغرض على وجه التحديد.

ولأغراض البحث الموزون يتم تحديد أوزان للمصطلحات عند صياغة استفسار البحث، للدلالة على أهميتها النسبية، أو مدى تركيز المستفيد عليها. ومن الممكن إعطاء الأوزان في شكل رموز (كرمز النجمة * في نظام إرك **ERIC**(*) الذي يدل على الواصفات الرئيسة، على سبيل المثال) أو رقمياً (عشرياً وأرقام صحيحة). ويتم تحديد مدى الوزن (كمدى الوزن من ١ إلى ٥، على سبيل المثال، حيث ٥ هي أعلى نقطة) على نحو يختلف من نظام إلى آخر. ويشترط بالطبع لإجراء البحث الموزون أن يكون الوزن قد طبق في مرحلة التنظيم، وبشكل أكثر تحديداً في مرحلة التكشيف. فمن الممكن على سبيل المثال في تطبيق البحث الموزون على الاستفسار "التصفية ٣ والجدال ٦" **filtering 3 AND controversy** أن يتوقع المستفيد الحصول على الوثائق التي تم فيها تسجيل هذين المصطلحين باستعمال مقياس الوزن نفسه؛ أي تحديد الوزنين ٣ و٦ للمصطلحين "التصفية" و"الجدال" على التوالي.

ومن الممكن استعمال العتبة **threshold**، وهي مفهوم مرتبط ارتباطاً وثيقاً بالبحث الموزون، لتحديد متى تلبي نتائج عملية البحث شروط الاستفسار الموزون. فإذا افترضنا أن كانت العتبة بالنسبة لعملية البحث "التصفية ٣ والجدال ٦"

(*)مرصد بيانات متخصص في التربية وما يتصل بها. (المترجم)

filtering 3 AND controversy 6، قد وضعت عند ٩، حينئذ لا يمكن النظر لأي نتائج بعتبة أقل من ٩ (كوثيقة في الموضوع نفسه ولكن بـوزن ٣ بالنـسبة "للتـصفية" و٥ بالنسبة للجدال على سبيل المثال) بوصفها صالحة في هذه الحالة.

ومما لا شك فيه أن إقرار الـوزن ليـتم تحديـده للمـصطلحات، أمـر حاسـم في البحث الموزون. وهناك خوارزميات وزن مختلفة تستخدم فعـلاً؛ وموقـع المـصطلح، وتقارب المصطلح، وتردد المصطلح (**tf**)، والـتردد العكـسي للوثيقـة (**idf**)، وتقـدير الفرد، هي المعايير الأساس بالنسبة لتحديد الأوزان. إلا أن مـن بـين تلـك المعـايير، وخصوصاً تقدير الفرد، أو الطريقـة الحتميـة أو القطعيـة **deterministic** **(Davis, 1997)** ما يمكن تطبيقه عملياً، للبحث الموزون، نظراً لأن المستفيد حين قـام بتحديـد الأوزان لمصطلحات الاستفسار، لم يكن على دراية بالوثائق التي يمكـن اسـترجاعها. إلا أن جميـع المعـايير الأخـرى الخاصـة بتحديـد الأوزان، تتوقـف عـلى خـواص المصطلحات المقتبسة أو المقتطفة من الوثائق الأصـلية. ولهـذا، فـإن أسـاليب الـوزن القائمة على موقع المصطلح، والمعايير الأخرى لا تبدو قابلة للتطبيق إلا على التكشيف الموزون فقط.

ويبدو البحث الموزون القائم على القـرار الفـردي، ذاتيـا جـدا. إلا أن تطبيقـه يعتمد فعلاً أيضاً على الجمع بين عوامل كالحاجة إلى المعلومات، وطبيعة نظم استرجاع المعلومات، وأنواع النتائج المتوقعة. وبعبارة أخرى، فإن المستفيد هو من يقرر الأوزان بالنسبة لمصطلحات الاستفسار، واضعاً تلك العوامل المحددة في الحسبان، ولهذا فـإن تخصيص الأوزان في أثناء صياغة الاستفسار، ليس مجرد مهمة تعسفية.

وكما سبق أن بينا فإن هناك معايير لتحديد الأوزان، يمكن الاختيار مـن بينهـا، من أجل التكشيف الموزون، أكثر مما هو متاح بالنسبة للبحث الموزون. ومن بين تلـك المعايير، دلالة موقع المصطلح على مكان ورود المصطلح في الوثيقة. وكقاعدة لا محيـد

عنها، فإن المصطلح الذي يرد في أماكن كالعنوان، أو عناوين الأقسام، والجمل الموضوعية **topic sentences**، يمكن أن يحظى بوزن أعلى من تلك المصطلحات التي ترد في أقسام أخرى من الوثيقة. ويعني تقارب المصطلحات تلك المسافة الفاصلة بين مصطلحات التكشيف في الوثيقة. ويمكن القول بوجه عام، أنه كلما ضاقت المسافة، أو كلما ازداد التقارب بين المصطلحات موضوع النظر، ارتفع وزن التكشيف الذي يمكن أن تحظى به المصطلحات. ومن ناحية أخرى، فإن تردد المصطلح (ت م **tf**) يُعرَّف بوصفه عدد المرات التي يرد بها المصطلح في وثيقة ما. وكما أوضحنا من قبل، فإن المصطلحات عالية التردد فقط، ليست بالضرورة مصطلحات تكشيف مرشحة جيدة، لأنها يمكن أن تكون كلمات وظيفية أو نحوية، أو كلمات لا تحمل معنى مهما في الاسترجاع. وفي مقابل ذلك، فإن أفضل مصطلحات كشفية، أو تلك المصطلحات التي يمكن أن تحظى بأوزان عالية، هي تلك التي ترد بكثافة في وثائق بعينها، ونادراً ما ترد في بقية وثائق مرصد البيانات (**Salton, 1989**). ومن ثم فإن عدد الوثائق التي يخصص لها مصطلح معين، ينبغي أن يوضع فى الحسبان عند إقرار الأوزان. ويعرف هذا الأسلوب بالتردد العكسي للوثائق **invesse document frequency (idf)**، نظراً لأن الأوزان التي يتم تحديدها تتناسب عكسيا مع تردد الوثائق. ويعني ذلك على وجه التحديد، أنه كلما قلت الوثائق التي يخصص أو يعين لها المصطلح، ارتفع الوزن الذي تحظى به.

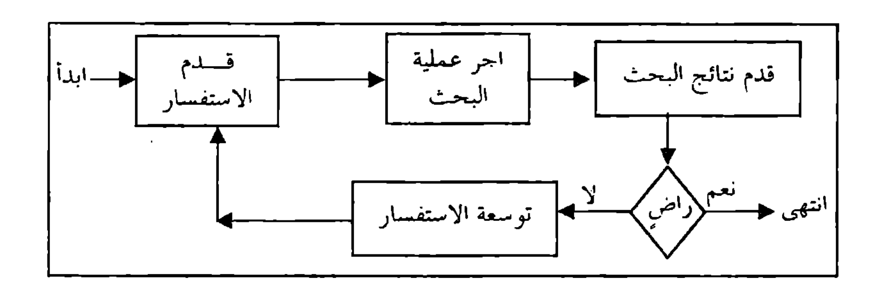
ولما كان تردد المصطلحات غالباً ما يستعمل مرتبطاً بالتردد العكسي للوثائق، لتحديد الوزن الذي يمكن أن تحظى به المصطلحات، فإن هذا الأسلوب يسمى أسلوب الوزن بناء على تردد المصطلحات، والتردد العكسي للوثائق. وطول الوثيقة (ط و **dl**) أحياناً ما يوضع أيضاً في الحسبان، جنباً إلى جنب مع أسلوب الوزن، بناء

على تردد المصطلحات والتردد العكسي للوثائق، وذلك لإضافة معيار آخر في إقرار أوزان المصطلحات. وإذا ما ظل كل من تردد المصطلح والتردد العكسي للوثائق ثابتين، فإنه كلما طالت الوثيقة التي تشتمل على المصطلح، ارتفع الوزن الذي يمكن أن يحظى به المصطلح. وقد اتبع أسلوب الوزن بناء على تردد المصطلحات والتردد العكسي للوثائق وطول الوثيقة، على نطاق واسع، في مشروع مؤتمرات استرجاع النصوص (ترِك TREC) Text REtrieval Conference (Spärck Jones, 2000) ونقدم مناقشة تفصيلية لسلسلة ترِك TREC في الفصل الحادي عشر.

وتستخدم أساليب وزن أخرى، كأساليب الاحتمالات، في التطبيقات العملية أيضاً. وفضلاً عن ذلك، فإن أساليب الوزن هذه، غالباً ما تطبق في تجمعات مختلفة مع طرق أخرى؛ فمن الممكن على سبيل المثال، تطبيق أسلوب موقع المصطلح جنباً إلى جنب مع تقارب المصطلحات في إحدى خوارزميات الوزن. ويمهد البحث الموزون سبيلاً آخر للاسترجاع الفعال للمعلومات.

توسعة الاستفسارات:

حتى وإن حظيت صياغة الاستفسار بالتفكير العميق، وهذا موضوع نستكشفه بمزيد من التفصيل في القسم ٥/ ٣، فإنه لا يمكن للمستفيد أن يضمن عدم حاجة استفسار البحث إلى المراجعة أو التوسعة، بمجرد الانتهاء من صياغته. وتوسعة الاستفسارات أسلوب للاسترجاع، يكفل للمستفيد النهائي فرصة الارتفاع بمستوى الأداء، بمراجعة استفسارات البحث بناء على النتائج التي تسترجع فعلاً. وعمليات البحث اعتماداً على توسعة الاستفسار تكرارية وتفاعلية، كما يتضح في الشكل رقم ٥/ ١.



الشكل رقم ٥/ ١ عملية توسعة الاستفسارات

ويتم فحص النتائج التي يتم الحصول عليها من عمليات البحث، بعـد تقـديم كل استفسار، لتبين ما إذا كان من الممكـن التقـاط أي معلومـات مـن هـذه النتـائج، وتضمينها في صياغة جديدة للاستفسار. ومن الممكـن نظريـاً تكـرار عمليـة توسـعة الاستفسارات إلى أن يتم الحصول عـلى نتـائج مرضـية. ويمكـن للتفاعـل بـين نظـام استرجاع المعلومات والمستفيد، أثناء عملية توسـعة الاستفـسارات، أن يـساعد عـلى الارتفاع بمستوى أداء الاسترجاع.

وكما ذهب رسنك وفوهان **Resnick and Vaughan, (2006)**، فإن توسعة الاستفسارات إحدى الطرق التي تتبـع عمومـاً في الـسياق. والأسـلوب الآخـر هـو تقلـيص أو انكـماش الاستفـسارات. وتنطـوي توسـعة الاستفـسارات عـلى إضـافة مترادفات أو مصطلحات متصلة أخرى إلى الاستفسارات بمعامل "أو"، لزيادة عدد حالات المضاهاة التـي يـتم تحديـد معالمهـا، بيـنما يهـدف تقلـيص الاستفـسارات إلى التخلص من غموض أي مصطلحات متعـددة المعـاني، مـن أجـل ضـمان اسـترجاع حالات المضاهاة المناسبة دون سواها. ومن الممكن تقسيم توسـعة الاستفـسارات إلى ثلاث فئات، بناء على المصادر التي يتم منها اختيـار المـصطلحات المتـصلة مـن أجـل التوسـعة **(Gauch, Wang, Rachakonda, 1999)**. وعـلى وجـه التحديـد، فـإن

التوسعة اعتماداً على مصطلحات منتقاة من مجموعة فرعية من الوثائق المسترجعة، بناء على استفسار بعينه، تسمى التوسعة بناء على استفسار معين **query-specific**. وإذا ما تمت التوسعة بمصطلحات أمكن التحقق منها عن طريق تحليل محتويات مرصد بعينه للنصوص الكاملة، فإنها هنا توسعة بناء على رصيد معين **corpus-specific**. وإذا ما تمت التوسعة بمصطلحات واردة في مكانز على الخط المباشر، لم يصمم خصيصا لأي مجموعة بعينها، فإنها تسمى التوسعة بناء على لغة معينة **language-specific**.

ومن الممكن تنفيذ توسعة الاستفسارات يدويا أو آليا. وبالنسبة للتوسعة اليدوية للاستفسارات، يقرر المستفيد كيف يمكن الإفادة من نتائج عملية البحث، لكي يعدل الاستفسار. وبالنسبة للتوسعة الآلية للاستفسار (التي تعرف أيضاً بالتلقيم المرتد للصلاحية **relevance feedback**) فإنه يفترض أن تكون الوثائق المتربعة على القمة في نتائج عملية البحث، صالحة، ومن ثم تراعى في مراجعة الاستفسار. ولا يشارك المستفيد النهائي على نحو مباشر في عملية تعديل الاستفسار **(Grossman & Frieder, 1998; Salton, 1970)**. وعند هذه النقطة يمكننا أن نلاحظ أن توسعة الاستفسارات، ليست بالتعبير بالغ الدقة للدلالة على هذه العملية، نظراً لأن الاستفسار لن يتسع بالضرورة، وإنما يتعرض للتعديل دائماً. ومن ثم فإن تعديل الاستفسارات يبدو مصطلحاً أكثر ملاءمة بالنسبة لهذه العملية.

وهناك أسلوب آخر لتعديل الاستفسارات، يحظى أيضاً بالتأييد، ويتم تطبيقه فعلاً من جانب بعض نظم استرجاع المعلومات؛ ففي هذا الأسلوب ترد المصطلحات آلياً في قائمة خيارات **menu** منسدلة، كاقتراحات، بينما يقوم المستفيد بإدخال الاستفسار في صندوق البحث. وعلى الرغم من أن بعض المستفيدين قد يجدون هذا الأسلوب غير مشجع، يدعي آخرون (مثل **White & Marchionini, 2006**) أنه من الممكن لمثل هذا الأسلوب أن يدعم فعلاً التوسعة الفورية للاستفسارات.

وترتيب المخرجات طبقياً، الذي يتم في كثير من نظم استرجاع المعلومات، أسلوب مهم أساس بالنسبة للتلقيم المرتد للصلاحية، أو توسعة الاستفسارات آلياً. وعلى نحو مماثل لأساليب الوزن التي سبقت مناقشتها، فإن خوارزميات الترتيب الطبقي تستند، في أغلب الأحيان على معايير من قبيل موقع المصطلح، وتقارب المصطلح، وتردد المصطلح. وعادة ما يطبق كل نظام لاسترجاع المعلومات خوارزميات الترتيب الطبقي الخاصة به، باختيار معايير معينة. ومثل خوارزميات الترتيب الطبقي هذه، لا تنشر أو تتاح على الملأ عادة. وهناك معيار ترتيب طبقي جديد، يسمى أسلوب الرابطة الراجعة **backlink**، أدخله عام ١٩٩٨ جوجل **Google**، أحد نظم استرجاع معلومات الإنترنت، لترتيب نتائج البحث في العنكبوتية العالمية طبقياً (**Vidmar, 1999**). ويحكم أسلوب الرابطة الراجعة على جودة صفحة العنكبوتية، بإحصاء عدد مواقع العنكبوتية الأخرى التي توفر روابط فائقة **hyperlinks** بهذه الصفحة، وذلك من بين أشياء أخرى. وهناك المزيد من مناقشات أسلوب الرابطة الراجعة، في الفصل الثامن "نظم استرجاع المعلومات".

والاستفسار بالمثال **query by example** أسلوب آخر لتوسعة الاستفسارات، حيث يدل مصطلح "مثال" على النتائج التي استرجعت فعلاً؛ وبعبارة أخرى، فإن النتائج التي يتم استرجاعها، تعامل بوصفها أمثلة للمزيد من الاستفسارات. وفي نظم استرجاع الأصوات والصور، يمكن للمثال أن يأتي مباشرة من المستفيد، كالمخطط المرسوم باليد لزهرة، لا من المعلومات المقتطفة من نتائج عملية البحث. وعند البحث في الإنترنت، يمكن للمستفيد التعبير عن تلقيم الصلاحية المرتد، بالنقر ببساطة على الروابط الفائقة مثل "المزيد مثل هذا أو المزيد من هذا القبيل **More like this**".

وغالباً ما تدعو الحاجة إلى تعديل الاستفسارات في استرجاع المعلومات، لأنه يمكن لهذا التعديل أن يرتفع بمستوى أداء الاسترجاع، وذلك بمراجعة استفسار عملية البحث استناداً إلى المصطلحات المقترحة في صندوق عملية البحث، أو النتائج التي تم الحصول عليها فعلاً. وكأساليب للاسترجاع، فإن توسعة الاستفسارات وتلقيم الصلاحية المرتد على وجه الخصوص، يتمتعان بالتطبيق على نطاق واسع، فضلاً عن الاحتمالات المستقبلية العريضة في استرجاع المعلومات في العصر الرقمي.

البحث في مراصد بيانات متعددة:

يعني البحث في مراصد بيانات متعددة، الذي يعرف أيضاً بالبحث الموحد **federated**، البحث تزامنيًا في أكثر من نظام واحد لاسترجاع المعلومات. وينظر إلى مصطلح مرصد البيانات هنا عادة بوصفه مرادفاً للنظام(*). والحاجة إلى البحث في مراصد بيانات متعددة، ثلاثية الأبعاد؛ فمن الممكن، أولاً، البحث في نظام واحد لاسترجاع المعلومات، ألا يسفر عن ما يبحث عنه المستفيد، لأن لكل نظام لاسترجاع المعلومات حدود التغطية الخاصة به، والتوجه والمقومات أو الخواص. وفي هذه الحالة ينبغي اللجوء إلى نظم أخرى لاسترجاع المعلومات من أجل عملية البحث. ثانياً، يمكن للبحث في مراصد بيانات متعددة أن يكون بمثابة أداة اختيار، إذا لم يكن المستفيد متيقناً من أي النظم يمكن أن يكون الخيار الأفضل بالنسبة لاستفسار بعينه. ثالثاً، يمكن للنتائج التي يتم الحصول عليها من البحث في مراصد بيانات متعددة أن توحي أيضاً أو تدل على النظم الملائمة للمستفيد، لإجراء المزيد من عمليات البحث فيها.

(*) قد لا نتفق مع المؤلفة في ذلك، لأن مرصد البيانات database يمكن أن يكون أحد مكونات نظام استرجاع المعلومات. وبعبارة أخرى، فإن نظام الاسترجاع لا يقتصر على مرصد البيانات. (المترجم)

وعند البحث في مراصد بيانات متعددة، على المرء أن ينتبه، على وجه الخصوص لأوجه الاختلاف في نَظْم **syntax** الاستفسارات، واللغة، وإمكانات البحث التي يكفلها كل نظام؛ فمن الممكن لخاصية مألوفة وأساس في نظام ما، أن لا تحظى بالمساندة في نظم أخرى. وفضلاً عن ذلك، فإنه يمكن للخاصية نفسها التي تحظى بالدعم في كل النظم، ألا يتم التعبير عنها بالطريقة نفسها. فمن الممكن، على سبيل المثال، للمعامل البوليني و**AND** أن يستخدم في نظام ما، بينما يستخدم نظام آخر علامة زائد. كذلك يمكن للغات المستخدمة في النظم المتعددة أن تكون مختلفة أيضاً. وتشمل الاحتمالات، اللغة الطبيعية في مقابل اللغة المقيدة، واللغات المقيدة المختلفة، واللغة الطبيعية المستعملة في مجالات تخصصية مختلفة.

وهناك تحد آخر يواجهه البحث في مراصد بيانات متعددة، وهو كيفية معاملة البيانات مختلفة الصيغ **formats** (مثل آسكي **ASCII**، والفهرسة المقروءة آلياً [مارك **MARC**])، المختزنة في مراصد بيانات متعددة. وزد ٣٩.٥٠ **Z39.50** مواصفة معيارية للتعريف بخدمة استرجاع المعلومات، وتحديد مواصفات البروتوكول **Protocol Specification** بالنسبة لتطبيقات المكتبات، وضعت لمعالجة البيانات بجميع الصيغ لأغراض الاسترجاع، وذلك من بين أشياء أخرى. ومن ثم فإن نظم استرجاع المعلومات الملتزمة بالمواصفة المعيارية زد ٣٩.٥٠، يُمكن البحث فيها تزامنياً، بصرف النظر عن أوجه الاختلاف في الصيغ والتقارب الجغرافي و **(Michael & Hinnebusch, 1995)**. ولما كانت قد أقرت من جانب المعهد الوطني الأمريكي للمواصفات المعيارية (آنسي **ANSI**) **American National Standard Institute**، والمنظمة الوطنية للمواصفات المعيارية للمعلومات (نيزو **NISO**) **National Information Standards Organization**، فإن زد ٣٩.٥٠ تطبق على نطاق واسع في الفهارس المتاحة للجمهور على الخط المباشر، اعتماداً على العنكبوتية العالمية Web-based OPAC, s (وبباكس **WebPacs**)، وبعض الأنواع

الأخرى من نظم استرجاع المعلومات. واعتماداً على زد ٣٩.٥٠ يمكـن للمـستفيد أن يسترجع أيضاً المعلومات باستخدام واجهة تعامـل **interface** موحـدة الـشكل، أيـاً كانت واجهة التعامل التي توافرت للنظام في الأصل.

وانطلاقاً مـن زد ٣٩.٥٠ واسـتجابة للتوقعـات التقنيـة الراهنـة، فـإن خدمـة العنكبوتيـة للبحـث والاسـترجاع **Search/ Retrieve Web Service (SRW)**، والبحث والاسترجاع عن طريق المحدد الموحـد لمكـان المـصدر **Search/Retrieve via URL (SRU)**، بروتوكولان قصد بهما تحديد معالم شكل معياري لاستفـسارات البحث في الإنترنت، وكذلك بنيان الاستجابات. فعندما يقدم المستفيدون الاستفسار نفسه لعدة نظم لاسترجاع المعلومات، فإن النَظْم **syntax** الذي يعبر عن الاستفـسار يمكن أن يكون مختلفاً، وكذلك الحال أيضا بالنسبة لبنيان الاستجابات؛ إذ تشتمل كل استجابة لا على نتائج عملية البحث فحسب، وإنـما عـلى معلومـات الـصياغة أيـضاً. ويعالج كل من خدمة العنكبوتية للبحث والاسترجاع **SRW**، والبحث والاسترجاع عن طريق المحدد الموحد لمكان المصدر **SRU**، اللـذان يـسميان إختـصاراً **SRW/U**، هذه المشكلات، وذلك بتحديـد مواصـفات الـنَظْم **syntax** بالنـسبة للاستفـسارات والنتائج؛ فهذان الأسلوبان **SRW/U** يكفـلان، عـلى وجـه التحديـد، للمـستفيدين، ومتعهدي نظم استرجاع المعلومات، إمكانية توجيـه الاستفـسارات لمراصـد بيانـات الإنترنت، وإعادة النتائج المسترجعة بسلاسة، دون الحاجة إلى بروتوكـولات مـا وراء البحث **metasearch protocols** المعقدة، كالمواصفة المعيارية زد ٣٩.٥٠ على سبيل المثال **(Morgan, 2004)**. فبالمقارنة بهذه الاخـيرة، فـإن **SRW/U** أقـل تعقـداً، وإن كانت تقابـل دلاليـاً المواصـفة المعياريـة زد ٣٩.٥٠ الكلاسـيكية **(Levan, 2003)**. وتعمل مكتبة الكونجرس على نشر **SRW/U** وما يتصل به مـن مواصـفات معياريـة **(Library of Congress, 2008)**.

وفضلاً عما سبق ذكره من مشكلات وتحديات، ترتبط بالبحث في مراصد بيانات متعددة، فإن سبيل دمج النتائج المسترجعة من عدة مراصد بيانات، يفرض تحدياً آخر، فالمخرجات المرتبة طبقياً، على سبيل المثال، في سبيلها لأن تصبح، وعلى نحو متزايد هي القاعدة في تقديم نتائج عمليات البحث. ولا تزال خوارزميات الترتيب الطبقي تختلف من نظام إلى آخر. ولا يمكن توقع أن يكون الترتيب الطبقي ٠.٩٠ لإحدى الوثائق المسترجعة من أحد النظم، هو نفسه الخاص بالترتيب الطبقي من نظام مختلف. ويبدو انصهار البيانات أو اندماجها، أي الجمع بين النتائج المسترجعة من نظم مختلفة، اعتمادا على طرق معينة، حلا محتملاً لهذه المشكلة. فقد اقترح بلكن **Belkin** وزملاؤه **(1995)**، على سبيل المثال، مجموعة من الأساليب الخاصة بدمج نتائج عدة استراتيجيات بحث، في عدة مراصد بيانات، لإنتاج أفضل استجابة ممكنة لأحد الاستفسارات. ومسار دمج مراصد البيانات **Database Merging Track**، في مشروع مؤتمرات استرجاع النصوص (ترِك **TREC**)، الذي نتناوله في الفصل الحادي عشر، مخصص أو مكرس للتصدي لمشكلة دمج البيانات الواردة من مصادر غير متجانسة، لتقديم مخرجات موحدة **(Voorhess & Harman, 2000)**.

وعادة ما يحظى البحث في مراصد بيانات متعددة بالدعم، من جانب مرافق أو متعهدي استرجاع المعلومات (مثل ديالوج **DIALOG**)، التي تدير مراصد بيانات متعددة، وليس نظاماً واحداً لاسترجاع المعلومات. وينهض ما يسمى ما وراء محركات البحث على الإنترنت، بوظيفة مماثلة بالنسبة لاسترجاع المعلومات من مصادر مختلفة على الإنترنت. ونتناول هذا الموضوع بمزيد من التفصيل في الفصل الثامن "نظم استرجاع المعلومات".

اختيار تقنيات الاسترجاع:

هناك، كما تبين في القسم السابق، تقنيات استرجاع مختلفة، يمكن للمستفيد النهائي أن يختار من بينها ما يناسب مهمة بحث معينة. وهناك كثير من العوامل التي ينبغي أن توضع في الحسبان عند اختيار تقنيات الاسترجاع. إلا أنه في هذا القسم تركز مناقشة اختيار تقنيات الاسترجاع، على وظائف هذه التقنيات وأدائها في الاسترجاع.

وظائف تقنيات الاسترجاع:

تعمل كل واحدة من تقنيات الاسترجاع بطريقة معينة، بحيث يمكن للمستفيد الاعتماد عليها لتحقيق أهداف البحث الملائمة؛ فاستعمال البتر، على سبيل المثال، يسترجع مختلف أشكال المصطلحات، إذا كانت تتقاسم قطاعات مشتركة في طرق هجائها. وفي مقابل ذلك يستوعب البحث المشوش أخطاء الهجاء وناتج التحقق من الأحرف ضوئياً **OCR**، إذا ما كان من الممكن أن تكون هناك بعض حالات المضاهاة بين الشكلين المختلفين للمصطلح. ولهذا، فإن أول سؤال ينبغي أن يطرحه المستفيد عند اختيار تقنية الاسترجاع، يتعلق بما يمكن لتقنية الاسترجاع (كالبحث الموزون على سبيل المثال) أن تنجزه. ولا يمكن للمستفيد أن يعرف ما إذا كانت تقنية الاسترجاع موضوع النظر مناسبة لاستفسار معين، إلا بعد الإجابة عن هذا الاستفسار.

أداء الاسترجاع:

يقاس أداء الاسترجاع تقليدياً، وفي معظم الأحيان، بالتحقيق **Precision** والاستدعاء **Recall**، على الرغم من أن هذين المقياسين يشكلان دائماً بؤرة مناقشات جدلية. ولأغراض التوضيح تركز المناقشة التالية على التحقيق والاستدعاء كمقياسين لأداء الاسترجاع، بينما ننظر في وجهات النظر المختلفة حولهما، في الفصل الحادي عشر "تقييم نظم استرجاع المعلومات".

ويعرَّف التحقيق بأنه نسبة عدد الوثائق الصالحة المسترجعة إلى مجموع الوثائق المسترجعة من نظام ما. ويقيس التحقيق قدرة نظام استرجاع المعلومات على التمييز؛ أي القدرة على فصل غير الصالح عن الصالح. فإذا تم استرجاع مئة وثيقة بعملية بحث ما، وحكم على ٣٥ منها بأنها صالحة، فإن معدل التحقيق بالنسبة لعملية البحث هذه يكون ٣٥ بالمئة. ويعرَّف الاستدعاء، من ناحية أخرى، بوصفه نسبة عدد الوثائق الصالحة المسترجعة إلى إجمالي عدد الوثائق الصالحة في نظام ما؛ فهو يقيس القدرة الاسترجاعية لنظام استرجاع المعلومات. فإذا افترضنا أن هناك مئة وثيقة صالحة، حول موضوع معين في نظام ما، وتم استرجاع ٤٥ منها، حينئذ يمكن لمعدل الاستدعاء لعملية البحث هذه أن يكون ٤٥ بالمئة. وبالنسبة لكل من التحقيق والاستدعاء، فإنه كلما ارتفعت النسبة المئوية كان أداء الاسترجاع أفضل. إلا أنه من المستحيل الحصول على نتائج بحث مرتفعة في كل من التحقيق والاستدعاء، نظراً لعلاقة التناسب العكسي بين المقياسين.

وفي حدود ما يهم أداء الاسترجاع، فإن جميع تقنيات الاسترجاع التي نوقشت في القسم ٥/ ١، يمكن تصنيفها بوصفها تؤدي إلى الارتفاع بمستوى التحقيق (كالمعامل البوليني و**AND**، والبحث الموزون) أو تؤدي إلى الارتفاع بمستوى الاستدعاء (كالمعامل البوليني أو **OR**، والبحث المشوش)(*). ومن ثم فإن قرار اختيار تقنيات الاسترجاع ينبغي أن يتم اتخاذه واضعين في الحسبان أداء الاسترجاع الذي يتوقعه المستفيد. فإذا كانت نتائج البحث عالية التحقيق هي المتوقعة، فإنه يتعين اختيار تقنيات الاسترجاع المقابلة، والعكس صحيح، أي أنه إذا كانت نتائج البحث عالية الاستدعاء هي المتوقعة، فإنه يتعين اختيار تقنيات الاسترجاع المقابلة.

(*) يمكن أن يضاف هنا البتر. (المترجم)

تقنيات الاسترجاع والارتفاع بمستوى التحقيق:

يؤدي المعامل البوليني (وAND)إلى الارتفاع بمستوى التحقيق، وذلك بالجمع بين مصطلحين أو أكثر في استفسار البحث، من أجل تحديد المضمون الدقيق لهذا الاستفسار. فإذا كانت هناك، على سبيل المثال، ثلاثة مصطلحات هي التصفية، والجدال، والعنكبوتية **filtering, controversy and Web**، في استفسار بحث ما، فإنه يمكن الحصول على نتائج دقيقة بجمع **"ANDing"** المصطلحات الثلاثة كلها معاً. وإذا ما تضمن استفسار البحث مصطلحين فقط، فإن النتائج يمكن أن تزيد في العدد، ولكنها تتراجع في التحقيق، نظراً لأن القيد على المصطلح الثالث قد رفع.

أما المعامل البوليني (فيما عدا **NOT**) فيرتفع بمستوى التحقيق باستبعاد المصطلحات غير المرغوب فيها من الاستفسار. فإذا افترضنا أن المستفيد يريد البحث عن وثائق حول جدال التصفية، فيما يتعلق بالمواد المتوافرة في العنكبوتية، ولكن ليس حول برامج التلفزة، فإنه من الممكن استعمال المعامل فيما عدا **NOT** لتحقيق ذلك الهدف.

ويؤدي البحث بمراعاة شكل الحرف الأول **Case-sensitive** إلى زيادة التحقيق، وذلك بالتمييز بين الكلمات التي تبدأ بحرف كبير **upper case**، وتلك التي تبدأ بحرف صغير **lower case**. ويوضح المصطلحان تارجت **target** و **Target**هذه النقطة؛ فإذا كان المستفيد لا يهتم إلا بالعثور على الوثائق حول متجر تارجت **Target store**، فإن البحث بمراعاة شكل الحرف الأول يمكن أن يسترجع ما يريده المستفيد على وجه الدقة. وما لم يكن نظام استرجاع المعلومات يدعم البحث بمراعاة شكل الحرف الأول، فإنه لا يمكن للمستفيد أن يجد سبيلاً للتمييز بين الشكلين الهجائيين للمصطلح. ومن ثم فإن الوثائق حول كل من متجر تارجت **Target** وهدف **target** يتم استرجاعها بوصفها نتائج مشروعة، ويتأثر معدل تحقيق عملية البحث سلباً.

ومعامل "مع" **WITH**، الخاص بالبحث بالتقارب، طريقة أيضاً للارتفاع بمستوى التحقيق، نظراً لأنه يحدد الترتيب الذي ترد به المصطلحات التي يربطها العامل "مع" في الوثيقة الأصلية أو بديلها. فإذا كانت صيغة البحث هي: "المعلومات مع التقنيات **information WITH technology**، فإن هذين المصطلحين يتعين أن يردا بالترتيب نفسه في نظام الاسترجاع. ويمكن لأي تجميعات أخرى للمصطلحات (معلومات التقنيات، والمعلومات والتقنيات **technology information** و **information and technology**، على سبيل المثال)، ألا تسفر عن نتائج دقيقة، وإنما عن شوشرة **noise** في عملية البحث. ويعمل المعامل س مع **nWITH** على النحو نفسه، الذي يعمل به المعامل مع **WITH**، فيما يتعلق بالارتفاع بمستوى التحقيق، إلا أنه يكفل تفاوتاً في مدى س من الكلمات بين المصطلحين اللذين يربط المعامل بينهما.

ويكفل البحث بالحقل دقة عملية البحث، وذلك بقصرها على حقل بعينه. فكل حقل مخصص للدلالة على إحدى الخصائص المميزة للوثائق. وإذا كان الاستفسار حول إحدى خصائص الوثيقة (المؤلف على سبيل المثال)، فإنه من شأن عملية البحث أن تكون أكثر دقة، إذا كانت تقتصر على ذلك الحقل، مما يمكن أن تكون عليه عندما نجري بحثاً شاملاً في جميع الحقول. ومن الممكن للحقول الأخرى أن تشتمل على اسم المؤلف لأسباب مختلفة، إلا أن هذه الحقول لم يقصد بها التعبير عن مسئولية تأليف الوثائق.

ويحدد البحث الموزون، وهو أسلوب آخر للارتفاع بمستوى التحقيق، أوزاناً لمصطلحات الاستفسار. ومادام الأمر كذلك، فإنه بمقدور المستفيد تحديد نقاط تركيز الاستفسار، وبذلك يحصل على نتائج تضاهي توقعاته على نحو أكثر دقة. ولنأخذ صيغة عملية البحث "التصفية والجدال" **filtering AND controversy**، كمثال

ثانية. فبإمكان المستفيد الإعراب عن التركيز على كل من المصطلحين بطريقة مختلفة، وذلك بتحديد وزنين مختلفين لهما. وإذا ما حظي مصطلح "الجدال" بالوزن الأعلى، حينئذ يكون تركيز الاستفسار على "التصفية" بوصفها "جدالا". وبعبارة أخرى، فإنه يمكن للمستفيد أن يكون أكثر دقة بالنسبة لاستفسار معين عندما يستعمل البحث الموزون. وقصارى القول، فإن المعاملين "و" و"فيما عدا" في البحث البوليني، والبحث بمراعاة شكل الحرف الأول من الكلمة، والمعامل "مع" في البحث بالتقارب، والبحث بالحقل، والبحث الموزون، ويمكن أن تستعمل كلها، بقدراتها المختلفة، للارتفاع بمستوى التحقيق في مهام الاسترجاع.

تقنيات الاسترجاع والارتفاع بمستوى الاستدعاء:

لا يشترط المعامل البوليني "أو"، كما سبق أن تناولناه، سوى ورود أي من المصطلحين اللذين يربط بينهما، في الوثيقة المسترجعة. ومن ثم فإن نتائج عملية البحث من المفترض أن تزيد في عددها، ومن ثم في الاستدعاء. فإذا كان الاستفسار هو "التصفية أو الجدال" **filtering OR controversy** على سبيل المثال، فإن عملية البحث تسفر عن نتائج مناسبة عندما يرد أي من المصطلحين "التصفية **filtering**" أو "الجدال **controversy**" في النظام. وتبدو القيود على عملية البحث بالمعامل "أو" أقل بالمقارنة بالمعاملين البوليني الآخرين "و" و"فيما عدا". وكلما قلت القيود في عملية البحث ارتفع الاستدعاء.

ويؤدي البتر إلى توسعة نتائج عملية البحث، بالسماح لجزء بعينه من المصطلح أن يتخذ أشكالا مختلفة، مادامت النتائج تتقاسم فيما بينها عنصرا واحدا مشتركا (وليكن جذع الكلمة على سبيل المثال). وفي هذه الحالة تدخل جميع أشكال جذع الكلمة في الحسبان في عملية البحث؛ فمن شأن استفسار للبحث عن "الدوريات الإلكترونية" **ejournal***، على سبيل المثال، أن يسترجع وثائق تشتمل على

ejournal، أو **ejournals** أو **ejournalism**، أو **ejournalist**، أو أي مصطلح آخر يتصدره المقطع **ejournal**. ومن ثم فإن الاستدعاء بالنسبة لعملية البحث هذه يمكن أن يكون ولاشك أعلى مما يمكن أن يكون عليه بالنسبة لعملية بحث لم يتم فيها بتر مصطلح الاستفسار. وعندما يتم بتر مصطلح عملية البحث (**ejournal** مثلاً) فإن نتائج البحث لن تشتمل إلا على تلك الوثائق التي ورد فيها المصطلح بهجائه الذي ورد في الاستفسار على وجه التحديد.

ويكفل المعامل "قريب" **NEAR** للبحث بالتقارب، الفرصة لاسترجاع المزيد من الوثائق الصالحة، بالسماح للمصطلحات التي يربط بينها المعامل لأن ترد بأي ترتيب في مرصد بيانات النظام. ومن ثم، فإن عملية بحث تتم بناء على "المعلومات" قريب من "التقنيات" **information NEAR technology** يمكن أن تسترجع الوثائق حول "تقنيات المعلومات" **information technology** أو "معلومات التقنيات" **technology information**، وبذلك تزيد احتمالات ارتفاع الاستدعاء. ويعمل المعامل "س قريب"، **nNEAR** على نحو مماثل لمعامل "قريب" **NEAR**، إلا أنه أوسع قدرة على الاستيعاب، وذلك بالسماح بوجود أوجه اختلاف في حدود مدى س من الكلمات بين المصطلحين اللذين يربط المعامل "س قريب" فيما بينهما.

ويمكن للبحث المشوش **fuzzy** أن يؤدي إلى الارتفاع بمستوى الاستدعاء، بالكشف عن الأخطاء الناشئة عن الهجاء غير الصحيح، وأخطاء نظام التحقق من الأحرف بصرياً **OCR**، والمصادر الأخرى للأخطاء، وتصويب هذه الأخطاء. فإذا كانت هناك وثيقة حول "الهاتف الخليوي" **cellular phone**، إلا أنه حدث خطأ في هجاء المصطلح فأصبح **Celluler phone**، فإنه يمكن للبحث باستعمال أسلوب التشويش، أن يظل قادراً على استرجاع الوثيقة، وبذلك يزداد الاستدعاء بالنسبة لعملية البحث هذه.

وتهدف توسعة الاستفسارات إلى استرجاع المزيد من النتائج الصالحة، وذلك بتعديل الاستفسار، بناء على الدفعة السابقة من نتائج البحث. ومن الممكن لعملية التعديل هذه أن تتكرر، إلى أن يتم الحصول على نتائج صالحة كافية. والاستدعاء، مقياس نسبة الوثائق الصالحة المسترجعة إلى كل ما يتوافر في النظام حول موضوع الاستفسار، هو العامل الذي يمكن الارتفاع بمستواه ولا شك، بتوسعة الاستفسارات. فمن الممكن، على سبيل المثال، للجولة الأولى من عملية البحث، أن تسترجع خمس وثائق فقط صالحة بالنسبة للاستفسار "أنموذج الفضاء المتجهي" **vector space model**. وبعد النظر في النتائج التي يتم الحصول عليها مباشرة، يضاف اسم سالتون (**Salton, G** مثلاً) إلى استفسار جديد للجولة الثانية لعملية البحث. فإذا ما تم استرجاع ست وثائق صالحة أخرى نتيجة لذلك، فقد ارتفع الاستدعاء على نحو ملحوظ.

ويكفل البحث في مراصد بيانات متعددة فرصة أخرى للارتقاء بالاستدعاء، نظراً للاعتماد على المزيد من مراصد البيانات في عملية البحث. وحينما يتم بحث استفسار في عدة مراصد للبيانات، فإن فرصة الحصول على عدد أكبر من النتائج الصالحة، تكون أكبر مما يمكن أن يسفر عنه إجراء البحث في نظام استرجاع معلومات واحد فقط. وفضلاً عن ذلك، فإنه من الممكن لكل نظام لاسترجاع المعلومات، أن يشتمل على كميات معينة وفئات منتقاة فقط من المعلومات. إلا أنه من الممكن لمجموعة من نظم استرجاع المعلومات أن تشتمل، ولا شك، على فئات متنوعة من المعلومات وبكميات أكبر مما يمكن أن يتوافر في نظام واحد.

وكما بينا، فإن كلا من معامل "أو" **OR** للبحث البوليني، والبتر، ومعامل "قريب" **NEAR** للبحث بالتقارب، والبحث المشوش، وتوسعة الاستفسارات، والبحث في مراصد بيانات متغددة، كل هذه يمكن أن تؤدي إلى توسعة عمليات

البحث بطريقة أو بأخرى. وعلى الرغم من أن المزيد من النتائج، لا يمكن بالضرورة أن يسفر عن استدعاء مرتفع ، فإنه يمكن أن تزيد احتمالات الحصول على المزيد من النتائج الصالحة بالنسبة لاستفسار معين. يضاف إلى ذلك، أنه بالإلمام بالطريقة التي تؤثر بها كل تقنية من تقنيات الاسترجاع، في معياري الاستدعاء والتحقيق بالنسبة لمهام البحث، يمكن للمستفيدين اتخاذ قرارات صائبة، في اختيار تقنيات الاسترجاع الملائمة لعمليات البحث الخاصة بهم.

التعبير عن الاستفسارات:

ينبغي الإعراب عن الحاجة إلى المعلومات، شفوياً أو تحريرياً، قبل أن يكون بالإمكان إجراء عملية البحث اللازمة لتلبية هذه الحاجة. وتسمى الحاجة إلى المعلومات التي يتم الإعراب عنها، بطلب البحث أو طلب الاسترجاع، الذي عادة ما يكون باللغة الطبيعية. ومن الممكن حينئذ تحويل طلب البحث إلى استفسار، باستعمال إمكانيات نظام استرجاع المعلومات، كقواعد صياغة الاستفسار، وإجراءات الاسترجاع، واللغة المقيدة إذا كان النظام يستخدم لغة مقيدة. وتسمى عملية التحويل هذه بالتعبير عن الاستفسار أو تمثيل الاستفسار، وهي أكثر العوامل أهمية في الاسترجاع (**Spärck Jones, 2000**).

الخطوات العامة:

التعبير عن الاستفسار عملية فكرية، عادة ما تتكون من الخطوات التالية:

١. تحليل مفاهيم طلب البحث، بتقسيم الطلب إلى مجموعات من المفاهيم أو الأوجه.

٢. التفكر في المترادفات، والمصطلحات الأعرض، والمصطلحات الأضيق الخاصة بالمفاهيم التي تم التحقق منها.

٣. ترجمة هذه المصطلحات إلى اللغة المقيدة، إذا كان النظام يستخدم لغة مقيدة.

٤. استعمال المنطق البوليني للتعبير عن علاقة البدل "OR" بين جميع المصطلحات المترادفة في كل مجموعة من المفاهيم، وجمع "AND" الأوجه المناسبة، ثم استبعاد "NOT" أي أوجه غير مرغوب فيها.

٥. استخدام تقنيات استرجاع أخرى حيثما تدعو الحاجة.

وعلى الرغم من أن هذه الخطوات ليست جامدة أو ملزمة، فإنها تشكل جوهر التعبير عن الاستفسارات. ومن الممكن أن تكون هناك بعض أوجه الاختلاف في الممارسات. وفيما يلي مناقشة تحليلية متأنية لعملية التعبير عن الاستفسارات، نظرا لما لها من دور حاسم في استرجاع المعلومات.

تحليل المفاهيم:

الخطوة الأولى في التعبير عن الاستفسار هي تقسيم طلب البحث إلى مجموعات من المفاهيم أو الأوجه؛ فطلب البحث، على سبيل المثال، هو العثور على الوثائق التي تعالج تنقية العنكبوتية العالمية بوصفها موضوعاً جدلياً، لا بوصفها من تقنيات المعلومات. ويتبين بعد التحليل أن هناك ثلاثة مفاهيم واضحة المعالم في طلب البحث، كما يتضح في الجدول رقم ٥/ ١.

الجدول رقم ٥/ ١ تحليل مفاهيم طلب بحث

المفهوم (١)	المفهوم (٢)	المفهوم (٣)
تقنية العنكبوتية	الجدل	تقنيات المعلومات

وقد تصادف في هذه الحالة، أن كانت المصطلحات التي يتضمنها الطلب هي نفسها المفاهيم التي ينطوي عليها الموقف. إلا أن هناك مواقف لا يكون فيها تطابق تام بين المصطلحات والمفاهيم، فمن الممكن على سبيل المثال، لطلب بحث أن يشتمل على مصطلحي "الحافلات" **buses** و"مترو الأنفاق" **subways**، إلا أن المفهوم الأساس يتعلق "بالنقل العام" **public transportation**. وعلى ذلك فإن "النقل العام" **public transportation**، وليس الحافلات **buses** و"مترو الأنفاق" **subways**، ينبغي أن يستعمل لأغراض التعبير هنا. يضاف إلى ذلك أن الأسماء أو العبارات الاسمية فقط هي التي ينبغي أن تستعمل للتعبير عن المفاهيم. أما الأفعال، إن وجدت في الطلب، فيتم التعبير عنها بالمعاملات البولينية. ولا ينبغي اختيار أجزاء الكلام الأخرى للتعبير عن المفاهيم التي أمكن التحقق منها في طلب البحث.

اختلاف المصطلحات:

للمصطلحات في غالب الأحيان مترادفات، ومصطلحات أعرض، ومصطلحات أضيق، فضلاً عن التنويعات الأخرى. وهدف هذه الخطوة في التعبير عن الاستفسار هو التفكر أو تدبر أكبر عدد من التنويعات، حسبما تقضي الضرورة، بالنسبة للمفاهيم التي سبق التحقق منها، حتى يمكن التعبير عن المفاهيم على نحو شامل. ويوضح الجدول رقم ٥/ ٢ التنويعات المحتملة بالنسبة للمفاهيم الثلاثة التي تم التحقق منها في الخطوة السابقة.

الجدول رقم ٥/ ٢ مجموعات المفاهيم الخاصة بالتعبير عن الاستفسارات

المفهوم (١)	المفهوم (٢)	المفهوم (٣)
تنقية العنكبوتية	الجدال	تقنيات المعلومات
تقنيات الشبكات	المناظرة	التنقية
تقنيات الحاسبات	النزاع	الرقابة
		تنقية العنكبوتية
		تنقية العنكبوتية العالمية
		الرقابة على العنكبوتية العالمية

وكما تبين في الجدول رقم ٥/ ٢، فإنه لم يتم تسجيل جميع تنويعات المصطلحات المحتملة، نظراً لأسباب كالذيوع أو الانتشار أو الألفة، وكذلك الإيجاز. فتحت المفهوم (١) على سبيل المثال، لم نورد "العنكبوتية العالمية" **World Wide Web** كشكل مختلف بالنسبة "للعنكبوتية" **Web**. ويتوقف القرار حول ما إذا كان من المتعين تسجيل مصطلح بعينه، على عدة معايير ذاتية **subjective**، كالاستدعاء المتوقع، ومدى التآلف مع الموضوع. وفي الوقت الذي لا يتم فيه تضمين جميع الأشكال المختلفة للمصطلح في الاستفسار النهائي، بالنسبة لعملية البحث، فإنه يمكن لتسجيل عدد من الاحتمالات، أن يساعد في اختيار مصطلحات الاستفسار المرشحة. ويمكن لعملية تسجيل الأشكال المختلفة للمصطلحات، أن توفر أيضاً مصطلحات مداخل لاختيار المصطلحات المناسبة من اللغة المقيدة، إذا كان النظام يستعمل إحدى هذه اللغات.

تحويل المصطلحات:

إذا كانت إحدى اللغات المقيدة تستخدم كلغة لاسترجاع المعلومات، فإن المصطلحات التي يتم التعبير عنها بلغة طبيعية في الخطوة السابقة، ينبغي أن تحول تبعاً لذلك إلى اللغة التي وقع عليها الاختيار. ومالم يكن الأمر كذلك، فإنه يمكن للمستفيد التجاوز عن هذه الخطوة في التعبير عن الاستفسار، والانتقال إلى الخطوة التالية. ويشترط التحويل التآلف مع اللغة المقيدة المستخدمة، كما يمكن أن يسلك أيا من الطرق التالية:

- المقابل الدقيق: أن يكون للمصطلح الوارد، مقابل دقيق في اللغة المقيدة. ويبدو ذلك أيسر خيار من بين جميع طرق التحويل.
- المترادفات أو المصطلحات المتصلة: ليس للمصطلح موضوع النظر سوى مترادفات أو مصطلحات متصلة فقط في اللغة المقيدة. وينبغي العمل على اختيار المصطلح الأقرب في معناه إلى المصطلح الذي يتم تحويله.
- المصطلحات الأعرض: ليس للمصطلح الوارد سوى مصطلحات أعرض منه فقط في اللغة المقيدة. ومن شأن دقة المصطلح الأصلي أن تتاثر سلباً نتيجة لهذا التحويل.
- المصطلحات الأضيق: ليس للمصطلح موضوع النظر سوى مصطلحات أضيق منه في اللغة المقيدة. ومن شأن مجال المصطلح الأصلي أن يتقلص من جراء هذا التحويل.
- المصطلحات غير الواردة: أن يكون المصطلح المراد تحويله اسم علم (اسم مؤسسة، أو اسم شخص، أو اسم مكان، على سبيل المثال) أو مصطلحاً جديداً لا نظير له في اللغة المقيدة. ويتعين في هذه الحالة إيجاد مصطلح جديد،

يسمى محدد الهوية أو المعرف **identifier**، لأغراض التحويل، إذا حدث أن كان المستفيد اختصاصي معلومات. ومن الممكن إلى حد بعيد في هذه الحالة، أن يكون المصطلح الجديد هو نفسه المصطلح المراد تحويله.

وفيما عدا طريق المقابل الدقيق، تشترط جميع طرق التحويل الأخرى تفسير المصطلحات. وتتحكم دقة التفسير سلفاً في دقة عملية التحويل.

استخدام المعاملات البولينية:

إذا افترضنا أن كانت جميع المصطلحات التي جمعت تحت كل مفهوم في الجدول رقم ٥/ ٢ في الشكل الصحيح، فإن الخطوة التالية هي استخدام المعاملات البولينية، لربط المصطلحات والمفاهيم التي سبق التحقق منها، وفقاً لما بينها من علاقات. وعلى الرغم من أنه من الممكن أن تكون هناك بعض مظاهر التفاوت في التطبيق، فإن هناك قاعدتين عامتين لاستعمال المعاملات البولينية:

١. ربط المصطلحات في مجموعة المفهوم نفسها (كجميع المصطلحات الواردة في عمود ما في الجدول رقم ٥/ ٢ مثلاً) بالمعامل "أو" **OR**.

٢. ربط كل مفهوم في طلب البحث (أي في الصف في الجدول رقم ٥/ ٢) بالمعامل "و" **AND** أو المعامل "فيما عدا" **NOT**، تبعاً للحاجة.

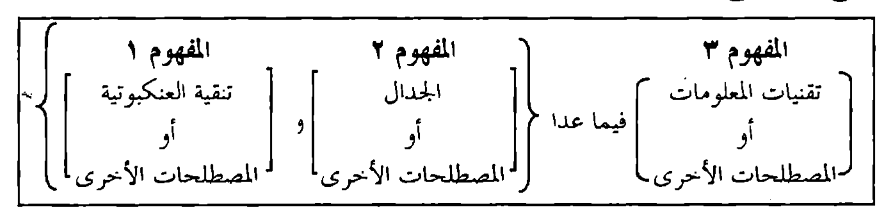
ويوضح الشكل رقم ٥/ ٢ هذه العملية رمزياً، بمجموعات المفاهيم التي وردت في الجدول رقم ٥/ ٢ بالنسبة لمثال طلب البحث.

ويبدو أن هناك نقطتين بحاجة إلى توضيح بشأن الشكل رقم ٥/ ٢؛ أولاهما، كم عدد المصطلحات من كل مجموعة من مجموعات المفاهيم، يمكن أن يتضمنها الاستفسار؟ وبعبارة أخرى، كم مصطلحاً ينبغي الربط بينها معاً بالمعامل أو **OR**، للتعبير عن كل مفهوم أمكن التحقق منه؟ وتتوقف الإجابة في الأساس، على عدد

النتائج التي يتوقع المستفيد الحصول عليها. وكلما زاد عدد المصطلحات المرتبطة معاً بالمعامل "أو" OR ازداد عدد النتائج التي يتلقاها المرء. وعلى النحو نفسه، فإنه كلما قل عدد المصطلحات المرتبطة معاً بالمعامل "أو" OR انخفض عدد النتائج التي يمكن أن يحصل عليها المرء. والأسلوب المألوف المتبع في اتخاذ مثل هذا القرار، هو وضع جميع المصطلحات الواردة مرتبة وفق أولويات، ثم اختيار المصطلحات القليلة الأولى الواردة في القائمة، لكي يتم ربطها معاً بالمعامل "أو" OR تبعاً لحاجة المرء. ولا نحدد في الشكل رقم ٥/ ٢ كم مصطلحاً يمكن أن تستعمل للتعبير عن المفهوم المقابل. وكل ما هنالك أنه يفضل أن نبين أن هناك مصطلحات أخرى متاحة للمعامل "أو" OR.

والنقطة الثانية في الشكل رقم ٥/ ٢ التي تحتاج إلى توضيح، هي استعمال الأقواس؛ ففي مناقشتنا للبحث البوليني، في القسم ٥/ ١/ ١/ ١، عرضنا ترتيب التجهيز بالنسبة لعمليات البحث البوليني المركبة. والمثال الموضح في الشكل رقم ٥/ ٢ عملية بحث بوليني مركبة، تتطلب استعمال الأقواس لتغيير الترتيب الطبيعي للتجهيز. وما لم تستعمل الأقواس فإنه يمكن لنتائج عملية البحث أن تكون أي شيء، إلا ما كان المستفيد يقصد من وراء عملية البحث. وإذا افترضنا أننا استبعدنا جميع الأقواس في الشكل رقم ٥/ ٢، واحتفظنا بجميع المصطلحات المقترحة في ذلك الشكل، وبالترتيب نفسه أيضاً، فإنه يمكن للمصطلح الأول الوارد تحت المفهوم ٣ (أي تقنيات المعلومات) أن يتم استبعاده من مجموعة النتائج التي تنشأ فيما يتعلق بالمصطلح الأخير الوارد تحت المفهوم ٢ (أي "النزاع" dispute). كذلك يمكن أن تنشأ نتائج أخرى، غير متوقعة في النهاية. ومن ثم فإنه ينبغي استخدام الأقواس إذا كان للترتيب الطبيعي للتجهيز أن يتغير.

وتفسر أهمية الأسلوب البوليني في استرجاع المعلومات نفسها بنفسها؛ إذ يتعين استخدام المنطق البوليني في التعبير عن الاستفسار، مالم تكن عملية البحث تشتمل على مصطلح واحد فقط. إلا أن تطبيق عملية البحث البوليني أكثر تعقداً بالفعل مما تبدو. وللمزيد من مناقشة تعقد البحث البوليني ومشكلاته، راجع الفصل السابع "نماذج استرجاع المعلومات".



الشكل رقم ٥/ ٢ تطبيق المعاملات البولينية

استعمال تقنيات الاسترجاع الأخرى:

هناك الكثير من التقنيات الأخرى التي يمكن للمستفيد اتباعها للتعبير عن الاستفسار بدقة ووضوح؛ فبالنسبة لعملية بحث "تنقية العنكبوتية **Web filtering**، على سبيل المثال (انظر الجدول رقم ٥/ ٢) يمكن للمستفيد النظر في:

- ما إذا كان البحث بمراعاة شكل الحرف الأول متاحاً للتمييز، حيث تدل **Web** على العنكبوتية العالمية **World Wide Web**، وتعني **web** نسيج العنكبوت **spider web**.
- ما إذا كان من الممكن بتر المصطلح **debate**.
- ما إذا كان من الممكن استعمال معاملات التقارب مثل "مع" **WITH** أو "قريب" **NEAR** للتعبير عن "تقنيات المعلومات" **information technology**.

- ما إذا كان من الممكن لعملية البحث أن تقتصر على حقول معينة، كحقل العنوان أو حقل الواصفات.
- ما إذا كان نظام استرجاع المعلومات الذي وقع عليه الاختيار يدعم البحث المشوش.
- ما إذا كان من الممكن تحديد أوزان لأي من المصطلحات التي وقع عليها الاختيار.
- ما إذا كان التلقيم المرتد الخاص بالصلاحية يحظى بدعم النظام، أو ما إذا كانت توسعة الاستفسار يمكن أن تتم يدوياً.
- ما إذا كان من الممكن البحث في مراصد بيانات متعددة.

وكما سبق أن بينا، فإن قليلاً من نظم استرجاع المعلومات، تكفل جميع إمكانيات البحث التي ناقشناها حتى الآن. ومن ثم فإن قائمة تقنيات الاسترجاع هذه إنما هي بمثابة قائمة مراجعة، وليست قائمة "ملزمة". وفي هذه المرحلة يمكن للاستفسار أن يكون قد تم التعبير عنه كاملاً، وأصبح جاهزاً لأن يقدم لإجراء عملية البحث.

والتعبير عن الاستفسار، بوصفه تعبيراً عن مواصفات المعلومات، عملية معقدة إلى حد بعيد. وعلى الرغم من أننا قد لخصنا هذه الخطوات الخمس، لتوضيح النقاط الجوهرية في التعبير عن الاستفسار، فإن التطبيق الفعلي قد لا يتكون من العدد نفسه من الخطوات، أو لا يتبع الترتيب نفسه. فالمستفيدون المتمرسون، على سبيل المثال، يميلون للجمع بين بعض الخطوات معاً، بينما يمكن للمستفيدين المبتدئين أن يكونوا عزوفين عن اتباع أي مسارات مختصرة.

وخلاصة القول، فإن الخطوة ١ في التعبير عن الاستفسارات، تتعلق بتحليل **parsing** طلبات البحث إلى مفاهيم، وتهتم الخطوتان ٢ و٣ بترجمة المفاهيم إلى مصطلحات الاستفسار. أما الخطوتان ٤ و٥ فتركزان على تطبيق مختلف أساليب الاسترجاع.

ولما كان المستفيد يختلف، وطلب البحث يختلف، ونظام استرجاع المعلومات يختلف، فإن التعبير عن الاستفسار ينبغي أن يبرز هذه الظاهرة، بوضع تفرد العملية أو اختلافها في الحسبان.

مشكلات التعبير عن الاستفسارات:

التعبير عن الاستفسارات، كما ناقشناه، ليس عملية ميكانيكية، وإنما عملية فكرية، تتطلب إعمال العقل والاجتهاد. يضاف إلى ذلك أن هذه العملية الفكرية تنطوي على عدة مشكلات.

ويعد تحليل المفاهيم أولى المشكلات في التعبير عن الاستفسارات؛ إذ ينبغي أن تتوافر للمستفيد المعرفة الكافية، والخبرة، والمهارات لكي يتحقق من المفاهيم التي يتضمنها طلب البحث، والتعبير عن هذه المفاهيم. ومن الممكن للذاتية أن تتسلل بسهولة إلى العملية. والتحليل غير الدقيق ليسن بالظاهرة غير المألوفة.

واللغة هي المشكلة الثانية في التعبير عن الاستفسارات؛ فاللغة الطبيعية ثرية، مرنة، غامضة، ومبهمة في بعض الأحيان، بينما اللغة المقيدة جامدة، ومصطنعة، ومن الصعب المحافظة على تجددها. إلا أن طلب البحث ينبغي التعبير عنه بأقصى درجات الدقة بكل من اللغتين. ويمكن لهذا التحويل أن يؤدي إلى بعض مظاهر التفاوت في عملية التعبير، ويؤثر في أداء الاسترجاع في النهاية. ويضاعف استعمال اللغة المقيدة من صعوبة التعبير عن الاستفسارات، نظراً لأن تآلف المستفيد مع هذه اللغة يتطلب وقتاً وجهداً. وفي مقابل ذلك فإن لاستعمال اللغة الطبيعية في التعبير عن الاستفسارات سلبياته كما بينا في الفصل الرابع.

ويثير استخدام تقنيات الاسترجاع المناسبة مشكلة أخرى في التعبير عن الاستفسارات؛ فكل نظام لاسترجاع المعلومات يحدد معالم مواصفاته وطريقة

استعماله فيما يتعلق بأساليب الاسترجاع (فعلامة زائد + على سبيل المثال، تعني في بعض نظم استرجاع الإنترنت المعامل البوليني "و" AND، بينما تستخدم في نظم أخرى رمزاً للوزن). كذلك يستنفد التمكن من استعمال هذه الأساليب وقتاً، كما يتطلب المهارسة.

ومن الممكن لكل هذه المشكلات أن تؤدي إلى افتقار التعبير عن طلبات البحث إلى الدقة، الأمر الذي يفضي في النهاية إلى أخطاء المضاهاة بين التعبير عن الاستفسار والتعبير عن مواصفات المعلومات. وللتغلب على هذه المشكلات يمكن تثقيف المستفيد، وتدريبه من ناحية، والسعي وراء اختراقات أو تطورات جوهرية في البحث العلمي حول التعبير الآلي عن الاستفسارات.

الأسلوب الآلي:

هناك رغبة ملحة في التعبير عن الاستفسارات آلياً، إذا ما كان بالإمكان فعلاً تحقيق ذلك. فالأسلوب الآلي، شأنه شأن التكشيف الآلي والطرق الآلية الأخرى لمعالجة النصوص، يتوقف إلى حد بعيد على خصائص من قبيل تردد الكلمات المفتاحية، والتقارب، والموقع. كذلك تطبق الخوارزميات القائمة على الاحتمالات، واللغويات الحاسبية، والذكاء الاصطناعي أيضاً في بعض الأحيان. إلا أنه على عكس التكشيف الآلي، الذي ينطوي على أنشطة ميكانيكية وفكرية في الوقت نفسه، فإن التعبير عن الاستفسارات آلياً لا يتضمن سوى مكون فكري فقط. ولما كانت الحاسبات لا تزال غير قادرة على التفكير كالإنسان، فإنه ليس من الصعب تصور المشكلات التي يمكن أن تثار، إذا ما ألقي على الحاسبات عبء المهمة الفكرية الخاصة االتعبير عن الاستفسارات.

وقد اختبر الباحثون في استرجاع المعلومات التعبير عن الاستفسارات آلياً، في الماضي. كذلك حظي الأسلوب الآلي باهتمام مكثف في مشروع مؤتمرات استرجاع

النصوص "ترِك" TREC. وقد تبين، على سبيل المثال، أن الاستفسارات التي تصاغ آلياً، من الموضوعات **topics** كانت تعمل كما تعمل الاستفسارات التي تصاغ يدوياً إن لم يكن أفضل، في المرحلة الأولى من مشروع ترِك ١ **TREC 1**. ولم يكن هناك تميز أو تحسن واضح نتيجة لإعداد الاستفسارات يدوياً **(Spärck Jones, 1995)**. وفي المرحلة الرابعة، ترِك ٤ **TREC 4**، كان هناك قليل من النتائج الآلية القابلة للمقارنة بالنتائج اليدوية، أو تضارعها، بالنسبة للاستفسارات القصيرة. إلا أن الاتجاه في ترِك ٥ وترِك ٦ يرجح الصياغة اليدوية للاستفسارات، التي تتم عادة للاستفسارات الطويلة، إلى حد يفوق أداء الصياغة الآلية للاستفسارات بوضوح، بالنسبة للاستفسارات الطويلة نفسها **(Spärck Jones, 2000)**. ويشتمل الفصل الحادي عشر على المزيد من المناقشات حول سلسلة ترِك **TREC**.

وعلى الرغم من أن نتائج سلسلة ترِك **TREC** ليست نهائية أو حاسمة، فإنها تظهر قصور الأسلوب الآلي، على ضوء الوضع الراهن للبحوث في هذا المجال. وهناك حاجة إلى المزيد من الجهد للارتفاع بمستوى الطريقة الآلية للتعبير عن الاستفسارات.

المراجع

Belkin, N. J., et al. (1995). Combining the evidence of multiple query representations for information retrieval. *Information Processing & Management*, 31, 431-448.

Davis, Charles H. (1997). From document retrieval to web browsing: Some universal concerns. *Journal of Information, Communication, and Library Science*, 3(3), 3-10.

Gauch, Susan, Wang, Jianying, and Rachakonda, Satya Mahesh. (1999). A corpus analysis approach for automatic query expansion and its extensions to multiple databases. *ACM Transactions on Information Systems*, 17(3), 250-269.

Grossman, David A., and Frieder, Ophir. (1998). *Information retrieval: Algorithms and heuristics*. Boston: Kluwer Academic Publishers.

Kowalski, Gerald. (1997). *Information retrieval systems: Theory and implementation*. Boston: Kluwer Academic Publishers.

LeVan, Ralph. (2003). Z39.50 as a web service. Retrieved December 1, 2008, fromstaff.oclc.org/-levan/docs/srw-niso20030430.ppt

Library of Congress. (2008). SRU: Search/Retrieval via URL. Retrieved December 1, 2008, from www.loc.gov/standards/sru

Michael, James J., and Hinnebusch, Mark. (1995). *From A to Z39.50:A networking primer*. Westport, CT: Mecklermedia.

Morgan, Eric Lease. (2004). An introduction to the Search/Retrieve URL Service (SRU). Ariadne, 40 (July 2004). Retrieved December 1, 2008, from www. ariadne.ac.uk/ issue40/ morgan

Resnick, Marc L, and Vaughan, Misha W. (2006). Best interface and future visions for search user interfaces. *Journal of the American Society for Information Science and Technology*, 57(6), 781-887.

Salton, Gerard. (1970). Automatic text analysis. *Science*, 168, 335-343.

Salton, Gerard. (1989). Automatic text processing: *The transformation, analysis, and retrieval of information by computer*. Reading, MA: Addison-Wesley.

Smith, Elizabeth S. (1993). On the shoulders of giants: From Boole to Shannon to Taube: The origins and development of computerized information from the mid-19th century to the present. *Information Technology and Libraries*, 12(2), 217-226.

Spärck Jones, Karen. (1995). Reflections on TREC. *Information Processing & Management*, 31(3), 191-314.

Spärck Jones, Karen. (2000). Further reflections on TREC. *Information Processing & Management*, 36(1), 37-85.

Vidmar, Dale J. (1999). Darwin on the web: The evolution of search tools. *Computers in Libraries*, 19, 22-28.

Voorhees, Ellen M., and Harman, Donna. (2000). Overview of the sixth Text REtrieval Conference (TREC-6). *Information Processing & Management*, 36(1), 3-35.

White, Ryen W., and Marchionini, Gary. (2006). Examining the effectiveness of real-time query expansion. *Information Processing & Management*, 43(3), 685-704.

الفصل السادس

طرق الاسترجاع

من الممكن تحقيق استرجاع المعلومات بالبحث أو التنقيب **searching**، أو بالتصفح **browsing**، أو بالطريقتين معا. ويتوقف اختيار طريقة الاسترجاع المناسبة على طبيعة حاجة المستفيد إلى المعلومات، وذلك من بين عوامل أخرى. ونحاول في هذا الفصل استكشاف الطرق الثلاث المختلفة للاسترجاع، إلى جانب خصائص هذه الطرق وتطبيقاتها.

وقد صاغ كل **Kall (2000)** التناظر الوظيفي التالي، عند مناقشة استرجاع المعلومات. فقد ذهب إلى أن العثور على إبرة (أي وثيقة أو معلومات) في كومة تبن **haystack** (أي نظام استرجاع المعلومات أو مرصد البيانات **database**) يمكن أن يعني:

١. إبرة معروفة في كومة تبن معروفة.

٢. إبرة معروفة في كومة تبن غير معروفة.

٣. إبرة غير معروفة في كومة تبن غير معروفة.

٤. أي إبرة في كومة ما.

٥. أمضى **sharpest** إبرة في كومة التبن.

٦. معظم أمضى الإبر في كومة التبن.

٧. جميع الإبر في كومة التبن.

٨. تأكيد غياب الإبر في كومة التبن.

٩. أشياء تشبه الإبر في كومة التبن .

١٠. دعني أعرف متى تظهر إبرة جديدة.

١١. أين أكوام التبن.

١٢. الإبر أو أكوام التبن - أياً كان.

وقد أمكن في القائمة السابقة حصر الاحتمالات المختلفة بشأن الإبرة وكومة التبن، إلا أن هذا الحصر لم يحقق الشمول. ومما لا شك فيه، أن البحث أو التنقيب هو الطريقة المناسبة لحالات معينة (رقم ١ على سبيل المثال)، ويبدو التصفح ملائماً لحالات أخرى (رقم ١٢ على سبيل المثال). ويمكن لبعض الحالات في القائمة (رقم ٥ على سبيل المثال) أن تتطلب الاستخدام المشترك لطريقتي الاسترجاع.

الاسترجاع بالبحث:

البحث أو التنقيب **searching** طريقة للاسترجاع، تستخدم منذ زمن طويل. وهو أيضاً موضوع دائماً ما يحظى بالمعالجة في استرجاع المعلومات، حيث يمكن أن يسمى البحث في مراصد البيانات، والبحث على الخط المباشر، والبحث في العنكبوتية العالمية، أو ما شابه ذلك. وبمجرد أن يتم التعبير عن الاستفسار، يكون المستفيد مستعداً لإجراء عملية البحث، من أجل استرجاع المعلومات من النظام.

خصائص البحث:

يقصد بالبحث بادئ ذي بدء العثور على ما يمكن أن يضاهي المصطلحات المحددة الواردة في الاستفسار، باتباع مختلف تقنيات الاسترجاع التي تناولناها في الفصل الخامس. ومن الممكن إجراء البحث عن طريق أي من نقاط الوصول **access**

points، الموضوعية أو غير الموضوعية. وعادة ما تشمل نقاط الوصول الموضوعية، في نظم استرجاع المعلومات، الواصفات **descriptors**، أو المعرفات أو محددات الهوية **identifiers**، أو وسيات أو رموز التصنيف، أو رؤوس الموضوعات، أو المراجع المستشهد بها، في البحث بالاستشهاد المرجعي، والمصطلحات الواردة في العناوين، والمستخلصات، أو النص نفسه، في البحث في النصوص الكاملة **(Hjørland, 1997)**. وتشمل نقاط الوصول غير الموضوعية المألوفة، اللغة، وسنة النشر، ونوع الوثيقة، وما شابه ذلك.

ويمكن للبحث أن يكون طريقة استرجاع فعالة، عندما يكون الاستفسار محدداً، والمستفيد على دراية على وجه اليقين بما هو بصدد البحث عنه. فإذا كان المستفيد يود، على سبيل المثال، العثور على كل بحوث سالتون **Salton**، التي نشرت في تسعينيات القرن العشرين، فإنه يمكن للبحث بحقل المؤلف وحقل عام النشر، أن يسفر عن النتائج المتوقعة(*). أما إذا كان المستفيد يود التعرف على سبيل المثال، على كل أولئك الذين سجلوا إسهامات في مجال استرجاع المعلومات، فإن البحث وحده لا يمكن أن يلبي هذه الحاجة.

وفي طريقة البحث، يمكن لتقنيات البحث كالمنطق البوليني، أن تكفل للمستفيد الجمع بين مختلف الأوجه في استفسار البحث، حيثما تدعو الحاجة. وباستثناء بعض نظم استرجاع الإنترنت، فإن معظم نظم استرجاع المعلومات تدعم إدخال المزيد من التعديل في استفسارات البحث، بمساعدة تقنيات الاسترجاع الملائمة (كالبحث بالحقل على سبيل المثال). وبإمكان النظام إنجاز كل هذه المهام، بمجرد أن تقدم له كعناصر في الاستفسارات. وهذه العملية مباشرة صريحة، ولا تلقي بعبء معرفي ثقيل على كاهل المستفيد.

(*) وذلك بالربط بين الوجهين بالمعامل البوليني "و" **AND** بالطبع. (المترجم)

غير أن البحث نشاط منضبط؛ وقلما تكون هناك فرصة للصدفة (أي اكتشاف شيء له قيمته فعلاً ولكنه لم يكن متوقعاً أثناء إجراء البحث)، ذلك لأن النظام لا يقدم سوى تلك النتائج التي تضاهي الاستفسار. يضاف إلى ذلك أن مهارات البحث لا تتحقق طبيعياً أو بالفطرة؛ إذ يتعين على المستفيد أن يتعلم كيف يبحث، وعليه أن يتمرس حتى يتمكن من إجراءات البحث. ويمكن لتكلفة التعلم والتمرس أن تكون عالية، إذا كان التعامل مع نظام استرجاع المعلومات يتم مقابل رسوم.

أنواع عمليات البحث:

من الممكن تقسيم عمليات البحث إلى أنواع مختلفة تبعاً لأغراض البحث؛ فعندما يكون لدى المستفيد وثيقة معينة في ذهنه، فإن البحث يسمى البحث عن وثيقة أو عن مادة. وعادة ما يتم إجراء البحث عن مادة معروفة بوجه عام، بناء على خواص وثيقة معروفة، كاسم المؤلف والعنوان. والمناظر الوثيق للبحث عن وثيقة معروفة هو ما ذهب إليه كل **Koll (2000)** "إبرة معروفة في كومة تبن معروفة".

وعندما يختار المرء موضوعات للبحث، فإنه يجري بحثا في الإنتاج الفكري لاكتشاف ما إذا كان هناك باحثون آخرون أنجزوا أي عمل في هذا المجال. ولا يفاجأ الباحثون إذا لم يعثروا على أي شيء وثيق الصلة باستفساراتهم، عند إجراء عمليات البحث، بل إنهم قد يسعدوا عندما يعرفوا أن لا شيء قد نشر بعد في الموضوع. ويسمى هذا الضرب من عمليات البحث بالبحث السلبي **(Stielow & Tibbo, 1998)**، أو كما عبر عنه كل **Koll (2000)** بأنه "التأكد من أن لا إبر في كومة التبن". وينبغي للبحث السلبي أن يكون شاملاً وضافياً، حتى يكون المستفيد النهائي واثقا من أنه ليست هناك وثائق تتصل باهتمامه أو يمكن أن تلبي حاجته(*).

(*) يلجأ المستفيد للبحث في الإنتاج الفكري عندما تتجاوز حاجته مجرد التحقق من وثيقة معينة. وللبحث في الإنتاج الفكري مساران؛ مسار راجع للتحقق من رصيد الإنتاج الفكري في مجال معين، في حدود لغوية وجغرافية وزمنية معينة، ومسار جارٍ لمتابعة ما يستجد من وثائق في موضوع معين. (المترجم)

ولمواكبة مجريات جبهة البحث، أو لأسباب أخرى مناظرة، يحرص المرء على اكتشاف ما هو جديد في مجاله، وذلك بالبحث في الإنتاج الفكري. ويسمى هذا الضرب من البحث بالبث الانتقائي للمعلومات (بام SDI). وكما بينا في الفصل الأول، فقد صمم لون **Luhn (1961)** طريقة البث الانتقائي للمعلومات، التي استخدمت على نطاق واسع، وخصوصاً في الأوساط المالية والعلمية. وبالنسبة لعمليات البحث لأغراض البث الانتقائي للمعلومات، يظل الاستفسار كما هو، إلا أنه يقدم لنظام استرجاع المعلومات دورياً. كذلك يمكن أيضاً تجهيز عمليات البحث للبث الانتقائي للمعلومات آلياً. وقد صور كل **Koll (2000)** عمليات البحث للبث الانتقائي للمعلومات بوصفها "دعني أعرف كلما ظهرت إبرة جديدة"(*).

ويدل البحث الموضوعي على ضرب آخر من عمليات البحث؛ فالمستفيد النهائي يشعر بالحاجة إلى المعلومات، ويود البحث عن بعض النتائج الصالحة. ولصياغة استفسار البحث، ينبغي استخدام نقاط الوصول الموضوعية التي أشرنا إليها في القسم ٦/ ١/ ١، وتلحق بها الخصائص غير الموضوعية (كاللغة وسنة النشر على سبيل المثال). وينبغي وضع تقنيات استرجاع مختلفة في الحسبان للحكم على معدل استدعاء وتحقيق عملية البحث. وقد عرض كل **Koll (2000)** عدة أمثلة للبحث الموضوعي، من منظور الاستدعاء والتحقيق عندما صاغ التناظرات التالية:

(*) يقوم البحث للبث الانتقائي للمعلومات على ثلاثة ركائز، وهي مرصد البيانات، وسمات المستفيد **user profile** ومضاهاة سمات المستفيد بأحدث قطاع في مرصد البيانات. وجدير بالذكر أن البث الانتقائي للمعلومات قد تطور عن الإحاطة الجارية **current awareness**.
راجع في ذلك:
حشمت قاسم: خدمات المعلومات، مقوماتها وأشكالها. القاهرة، دار غريب، ١٩٨٤.

- أي إبرة في كومة التبن: البحث منخفض الاستدعاء ولكنه مرتفع التحقيق.
- أمضى إبرة في كومة التبن: البحث مرتفع التحقيق.
- معظم أمضى الإبر في كومة التبن: البحث مرتفع الاستدعاء.
- جميع الإبر في كومة التبن: البحث في ذروة الاستدعاء.

وهناك إمكانات أخرى (عمليات بحث منخفضة التحقيق ولكنها مرتفعة الاستدعاء على سبيل المثال) فيما يتعلق بقياس ناتج الاسترجاع بناء على الاستدعاء والتحقيق. وبالمقارنة بأنواع عمليات البحث الأخرى، يبدو البحث الموضوعي أكثر الأنواع تعقداً، وذلك لأن النتائج التي يتم الحصول عليها من البحث، ينبغي تقييمها باستخدام المعيارين الجدليين، الاستدعاء والتحقيق. وكما سبق أن ذكرنا فإن الجدل حول التحقيق والاستدعاء كمقياسين لأداء الاسترجاع، سوف نتناوله بشيء من التركيز في الفصل الحادي عشر "تقييم تنظيم المعلومات واسترجاعها".

ويمثل بحث الفقرات **passage searching**، الذي ورد بإيجاز في القسم ١/ ٢/ ١ بوصفه استرجاع الفقرات، ضرباً آخر من عمليات البحث؛ إذ لا يتلقى المستفيد سوى الفقرة المناسبة فقط من الوثيقة، لا الوثيقة كاملة. ويتم الاعتماد بكثافة على مهمة التنقية في بحث الفقرات. وعلى الرغم من أنه لم يظهر إلا في السنوات الأخيرة، فإن بحث الفقرات ينطوي على احتمالات ضخمة بالنسبة لمساعدة المستفيد في الحد من عبء المعلومات الزائد، والارتفاع بمستوى التحقيق في الاسترجاع.

ويتناول هذا القسم خمسة أنواع مختلفة لعمليات البحث، وهي البحث عن مواد معروفة، والبحث السلبي، والبث الانتقائي للمعلومات، والبحث الموضوعي، وبحث الفقرات. وباتباع معيار آخر للتقسيم الفئوي، صنف بايزا - ييتس وريبرو - نتو **Baeza-Yates and Ribeiro-Neto (1999)** عمليات البحث إلى نوعين،

البحث المخصص أو الموجه **ad hoc** والتنقية **filtering**. وأيا كانت تسمية نوع عملية البحث، فإن الهدف هو تحديد أي أساليب البحث، وأي استراتيجيات البحث ينبغي اتباعها لتلبية احتياجات المستفيد.

استراتيجيات البحث:

استراتيجيات البحث هي السبل التي يسكلها المرء عند إجراء عمليات البحث. وفي العصر الذهبي للبحث على الخط المباشر، تحددت معالم الأنواع الأربعة التالية من استراتيجيات البحث، كما تم توضيحها **(Fenichel & Hogan, 1981)**.

طريقة كتل البناء:

تبدأ طريقة كتل البناء **building block** بعمليات البحث أحادية المفهوم **single-concept**. والمفاهيم الأحادية "كتنقية العنكبوتية والجدال" **web filtering and controversy**، الواردة في الجدول رقم ٥/ ١، هي تلك التي أمكن التحقق منها في تحليل المفاهيم. وبعد أن تكتمل جميع عمليات بحث المفاهيم الأحادية، يتم الجمع بين المجموعات الناتجة معاً باستعمال المعاملات البولينية الملائمة.

ويمكن لهذه الاستراتيجية أن تحلل مهمة البحث المركب، إلى مهام أبسط. كما أنها تكفل أيضاً للمستفيد إجراء التصويبات أو إدخال التعديلات، كلما دعت الحاجة، على نحو فوري أثناء إجراء عمليات البحث. فليس على المستفيد أن يعيد إجراء البحث كاملاً، نظراً لحدوث خطأ مطبعي بسيط أثناء إدخال صيغة البحث. ومن ثم فإن نهج كتل البناء يخفض الضغط على المستفيد، الذي يمكنه حينئذ التركيز أكثر على التفاعل مع النظام. ولهذه الأسباب يوصى بهذه الاستراتيجية بقوة لدى المستفيدين، وخصوصاً المبتدئين منهم.

طريقة الكرة الجليدية:

وتعني طريقة الكرة الجليدية **snowballing**، التي تعرف أيضاً بطريقة زراعة اللؤلؤ بالاقتباس **Citation pearl - growing (Fenichel & Hogan, 1981)**، أن نتائج البحث تزداد كماً، على نحو يشبه كثيراً ناتج دحرجة الكرة الجليدية في الثلج. ومما لا شك فيه، أن هذه الطريقة تهدف إلى الارتفاع بمستوى الاستدعاء. فباتباع هذه الاستراتيجية يجري المستفيد أولاً عملية البحث، ثم يعدل استفسار البحث، بناء على النتائج المسترجعة. وتنطوي عملية التعديل على النظر في النتائج، واختيار المصطلحات الصالحة (ككلمات العنوان والمواصفات مثلاً) ثم تضمينها في النص المنقح للاستفسار. ومن الممكن تكرار هذه العملية إلى أن يصبح المستفيد راضياً عن نتائج البحث تماماً.

فإذا ما استرجع بحث عن "الكتاب الإلكتروني" **electronic book** على سبيل المثال، وثائق تشتمل على مصطلحات مثل ستيفن كنج والكتاب الإلكتروني **Stephen King and ebook**، فإنه يمكن للمستفيد أن يتبع طريقة كرة الجليد، بتضمين هذين المصطلحين في صيغة البحث المنقحة، من أجل توسعة نتائج البحث. ونحن نعلم أن ستيفن كنج هو أول مؤلف شهير ينشر إلكترونيا(*)، وأن **ebook** شكل مختصر من **electronic book**. وطريقة كرة الجليد في جوهرها، مماثلة لأسلوب الاسترجاع المتطور، وهو توسعة الاستفسارات، الذي تناولناه في القسم ٥/ ١/ ٢/ ٣. وقد أطلق كورفاج **Korfhage (1997)** على هذه الطريقة اسم تجهيز الاقتباس **citation processing**. وكما تبين فعلاً، فإن طريقة الكرة الجليدية هذه مفيدة على وجه الخصوص، عندما تكون لدى المستفيد الرغبة في التحقق من المصطلحات المتصلة، من أجل توسعة عملية البحث.

(*) مؤلف روائي أمريكي، ألف خمسين رواية، فضلاً عن ثماني مجموعات قصصية، وبلغ إجمالي توزيع كتبه أكثر من ٣٥٠ مليون نسخة، وتحول بعضها إلى أفلام سينمائية ومسلسلات تلفزيونية. (المترجم)

طريقة التجزىء المتتابع:

تبدو طريقة التجـزىء المتتـابع **successive fraction** عكـس طريقـة الكـرة الجليدية؛ فعندما يبدأ البحث بمفهوم عريض، تتكفل طريقة البحث بالتجزىء المتتابع بتضييق البحث باستخدام مختلف أساليب التقييد أو التحديد. ويستخدم أحـد أنـواع التحديد المعاملات البولينية ومعامل التقارب، عند صياغة عبارة البحث. وكـما سـبق أن بينا، فإن معامل "فيما عدا" **NOT** يضيق البحـث باسـتبعاد مـصطلحات معينـة. ويحقق المعامل و**AND** الهدف نفسه بإضافة المزيـد مـن الـشروط في البحـث. كـذلك يضيق المعامل "مع" **WITH** البحث بتحديد مواضع الكلمات في استفسار البحـث. ومن الممكن تحقيق ضرب آخر من التضييق باستعمال خواص غير موضوعية كاللغـة، ونوع الوثيقة، وسنة النشر. ويبدو التضييق بمثل هذه الخواص غير الموضوعية أبـسط وأيسر من التضييق بالمعاملات.

لنفترض أنه قد تم إجـراء بحـث عـن "التنقيـة" **filtering**، وكانـت النتـائج تشمل كل شيء حول التنقية، حظى بالتغطية في نظام استرجاع المعلومات. فـإذا كـان هدفنا هو البحث عن معلومات حول تنقية العنكبوتية العالمية بوصفها جدالاً، وليس بوصفها من تقنيات المعلومات، فإننا يمكن أن نتبع طريقة التجـزىء المتتـابع لتحقيـق هدفنا؛ إذ يمكن على سبيل المثال، إضافة المصطلح "العنكبوتيـة العالميـة" **web** أولاً، إلى صيغة البحـث باسـتعمال المعامـل "و" **AND**. ومـن الممكـن إضـافة مـصطلح "الجدال" **controversy** بالطريقة نفسها. ومن الممكن بعد ذلـك اسـتخدام معامـل "فـيما عـدا" **NOT**، وذلـك لاسـتبعاد أي نتـائج تـشتمل عـلى مـصطلح "تقنيـات المعلومات" **information technology**. ومن الممكن تحقيق المزيـد مـن التـضييق، بقصر البحث على الوثائق التي نشرت قبل عام ١٩٩٦ على سبيل المثال.

وكما بينا، فإن طريقة التجزىء المتتابع تضيق البحث تدريجياً، باستعمال أساليب التحديد المتاحة في نظام استرجاع المعلومات. وتشترط استراتيجية البحث هذه أن يكون المستفيد متآلفاً مع مقومات التحديد المتوافرة في النظام، وأن يحافظ على تفاعله مع النظام أثناء البحث. وإذا ما وقع الاختيار على استراتيجية التجزىء المتتابع، فإنه يمكن للتكلفة أن تكون مهمة، بالنسبة للمستفيد من نظم استرجاع المعلومات على الخط المباشر (راجع القسم ٨/ ١) نظراً لأن هذه الاستراتيجية تتطلب وقت تجهيز أطول مما تتطلبه استراتيجيات البحث الأخرى. والواقع أن نظم استرجاع الإنترنت (راجع القسم ٨/ ٤) لا تفرض رسوماً على المستفيد، إلا أن مقومات التضييق التي توفرها مثل هذه النظم، لا تزال غير ملائمة إلى حد بعيد.

الأولوية لأكثر الأوجه تحديدا:

تطبق طريقة الأولوية لأكثر الأوجه تحديداً **most specific facet first** على عمليات البحث متعددة الأوجه أو المفاهيم (**Fenichell & Hogan, 1981**). وتفترض هذه الطريقة أن يكون المستفيد على دراية بجميع الأوجه، أو مجموعات المفاهيم التي تتضمنها عملية البحث، كما أن باستطاعته أيضاً التحقق من أكثر المفاهيم تحديداً. ويمكن لاستراتيجية البحث هذه أن تكون عالية الكفاءة؛ نظراً لأنها تبدأ عملية البحث بمصطلح يمكن أن يستغرق البحث عنه أقل قدر من الوقت. وفضلاً عن ذلك لا يكون هناك مبرر لمواصلة البحث، إذا كان عدد النتائج المسترجعة بالنسبة لأكثر الأوجه تحديداً، لا يمكن قبوله؛ فهناك على سبيل المثال ثلاثة أوجه في عبارة ''علاج الأورام الهرمونية العصبية في الرئة والتنبؤ بها'' **treatment and prognosis of neuroendocrine tumors in the lung**، وهي ''العلاج والتنبؤ''، و''الأورام الهرمونية العصبية''، و''الرئة''. ومن بين هذه الأوجه الثلاثة تبدو ''الأورام الهرمونية العصبية'' هي الوجه الأكثر تحديداً، وبذلك يتم البحث عنه

أولاً في هذه الطريقة. وإذا ما استرجع البحث عن "الأورام الهرمونية العصبية" وثيقتين فقط، فإنه لا يحتمل للبحث الذي يشمل الأوجه الثلاثة أن يسفر عن نتائج أكثر من الوثيقتين اللتين أمكن العثور عليهما فعلاً.

وفضلاً عن كفاءتها، يبدو تطبيق هذه الطريقة متطوراً جداً. فلما كان تحليل المفاهيم وحده مهمة شاقة، فإنه يصبح من المستحيل تقريباً، بالنسبة لمستفيد مبتدئ التحقق من أكثر الأوجه تحديداً، في طلب للبحث متعدد الأوجه. ومن ثم فإن طريقة الأولوية لأكثر الأوجه تحديداً لا يوصى بها بالنسبة للمبتدئين.

ويسمى أحد امتدادات طريقة الأولوية لأكثر الأوجه تحديداً، بطريقة الأولوية لثاني أكثر الأوجه تحديداً، التي تتعامل مع الوجه الذي يأتي تالياً من حيث التحديد، لإجراء عمليات البحث. إلا أن أياً من الطريقتين لا تستعمل غالباً في أوساط استرجاع المعلومات. وطريقة الأولوية لأقل الأوجه تردداً، التي تحظى بالمعالجة في **Fenichel and Hogan (1981)**، ببساطة، شكل مختلف من طريقة الأولوية لأكثر الأوجه تحديداً؛ إذ إن أكثر الأوجه تحديدا عادة ما يسفر عن أقل الأوجه ترددا(*).

في الطريق إلى طريقة سريعة وكفء:

كما ألمحنا في بداية هذا القسم، فإن كل ما تناولنا من استراتيجيات عمليات البحث قد اقترح وتطور في سياق ذروة البحث على الخط المباشر. فقد تغيرت بيئة الاسترجاع (كظهور نظم استرجاع الإنترنت على سبيل المثال) كما تغير من يقومون بالبحث أنفسهم (المستفيدون كباحثين دون مساعدة الوسطاء على سبيل المثال)، تغيرا

(*) واقع الأمر أن النظم الإلكترونية لاسترجاع المعلومات، وخاصة تلك التي تعمل على الخط المباشر، تبدأ عملية البحث والمضاهاة بأقل الأوجه ترددا، لأن مجموع ما يحتمل استرجاعه من وثائق، لا يزيد بحال عن عدد تلك الوثائق المرتبطة بذلك الوجه. (المترجم)

جوهريا منذ ذلك الوقت. وبينما لا تزال كل من طريقة كتل البناء وطريقة الكرة الجليدية، لهما الأفضلية من جانب بعض المستفيدين، فإن كثيرين يكتفون بمجرد طباعة كلمة مفتاحية واحدة يودون البحث بناء عليها، أو يدخلون عدة مصطلحات دفعة واحدة، تاركين مسافات بينها، ونادراً ما يفيدون من المعاملات البولينية وغيرها من مقومات البحث (Jansen, Spink & Saracevic, 2000; Siegfried, Bates, & Wilde, 1993; Wallace, 1993)، على سبيل المثال. ومن الممكن تسمية هذه الطريقة بالطريقة السريعة الكفء. ويمكن لقليل ممن يقومون بإجراء عمليات البحث اليوم، تجربة طريقة التجزيء المتتابع، أو طريقة البدء بأكثر الأوجه تحديداً. ومن ناحية أخرى، تدخل الآن بعض مقومات أو وسائل البحث (كاللغة، والمعاملات البولينية على سبيل المثال) في واجهات التعامل مع نظم استرجاع المعلومات. وبإمكان المستفيد ببساطة، اللجوء إلى هذه الخيارات من أشكال عمليات البحث، والأزرار المحددة سلفاً، أو قوائم الخيارات المنسدلة، دون طباعتها فعلاً في عبارات البحث.

الاسترجاع بالتصفح:

التصفح **browsing** شكل آخر رئيس لنشاط الاسترجاع. وبالمقارنة بالبحث، فإن التصفح لم يحظ بكثير من الاهتمام في أوساط استرجاع المعلومات حتى تسعينيات القرن العشرين. ونتيجة لتطور نظم الأسطوانات المكتنزة **CD-ROM**، ونظم الفهارس المتاحة للجمهور على الخط المباشر **OPAC**، والعنكبوتية العالمية فائقة التنظيم على وجه الخصوص، حظي التصفح بالذيوع بسرعة، وأصبح مجديا من الناحية الاقتصادية في استرجاع المعلومات، عندما لم يعد المستفيد بحاجة لأن يتحمل عبء رسوم الاتصالات، كما في حالة النظم التجارية لاسترجاع المعلومات على الخط المباشر؛ فبإمكان المستفيدين اليوم استرجاع المعلومات بالتصفح كما يحلو لهم.

ما التصفح؟

التصفح هو التماس المعلومات والتقاطها، بالإطلاع السريع والتمحيص وغير ذلك من الأنشطة. وفي استرجاع المعلومات يتجه المرء نحو التصفح:

- للحصول على المعلومات عندما لا يكون الموضوع محدداً بوضوح
- للعثور على المعلومات التي يصعب تحديدها صراحة
- لإلقاء نظرة عامة على المعلومات التي يوفرها نظام استرجاع المعلومات
- لاختيار المعلومات المناسبة من خليط من المواد الصالحة وغير الصالحة
- لاكتشاف المعلومات الجديدة واستيعابها
- عندما تحث بيئة استرجاع المعلومات على التصفح

ويمكن الاطلاع على أسباب إضافية للتصفح في **Marchionini (1995)** بالإضافة إلى معالجات ضافية. وفي حالات من قبيل "الإبر وأكوام التبن أيا كانت" **Koll (2000)**، يمكن للتصفح أن يكون طريقة استرجاع مفضلة.

وعند التصفح لا يكون المستفيد بحاجة لأن يعبر عن مشكلة الاسترجاع بمصطلحات محددة، كما يشترط في عملية البحث. وعلى هذا النحو فإن التصفح يتطلب جهداً فكرياً أقل مما يتطلبه البحث. إلا أنه بينما يقوم بالتصفح، فإن المرء ينبغي أن يظل متفاعلاً مع نظام استرجاع المعلومات، من أجل أن يفحص ويقيم المعلومات التي يتصفحها، ويصدر حينئذ أحكام الصلاحية **(Marchionini, 1995)**. ومن ثم فإن العبء المعرفي الملقى على عاتق المستفيد في التصفح أكبر مما هو عليه إذا ما اختار البحث نهجًا للاسترجاع.

وعلى عكس البحث، فإن التصفح يمكن أن يتم كطريقة للاسترجاع دون أي تدريب أو تمرس. وكما ذهب مارشيونيني **Marchionini (1995)**، فإن:

> [التصفح] طبيعي لأنه يربط بين الموارد البشرية العضلية والانفعالية والمعرفية، بالطريقة نفسها، التي يراقب بها البشر العالم الفيزيائي ويبحثون عن أشياء مادية. **(P. 100)**.

وبإمكان المرء أن يتصفح ما دام قادراً على التنفس. إلا أن هناك أساليب قليلة تطورت على النحو المناسب، لتيسير عملية التصفح. وفي متناول المستفيد بوجه عام توجيهات أو إرشادات قليلة، يمكن الاعتماد عليها، حول متى يواصل التصفح من أجل المزيد من المعلومات، ومتى يتوقف عن التصفح.

ويمكن للتصفح، من ناحية أخرى، أن يكون نشاطاً مجزياً؛ إذ يمكن أن يفضي إلى المصادفة، أي العثور على شيء مفيد على نحو غير متوقع، نظراً لأن التصفح يكفل للمستفيد البحث عن المعلومات عشوائياً وعلى نحو غير منظم. والاسترجاع بالتصفح طريقة لها مظاهر قوتها ومواطن ضعفها أيضاً.

أنواع التصفح:

على عكس البحث، فإن التصفح لا يبدو طريقة استرجاع منظمة. إلا أن هناك أنواعاً مختلفة من التصفح، أمكن التحقق منها ومعالجتها في الإنتاج الفكري (**Herner, 1970; Kowalski, 1997; Marchionini, 1995** على سبيل المثال). ويرى هيرنر **Herner (1970)** أن التصفح يمكن أن يكون موجهاً، أو شبه موجه، أو غير موجه، وذلك بالتحقق مما إذا كان المستفيد يلتمس كياناً محدداً، وتبعاً لإجراءات التصفح المتبعة. كذلك يصنف مارشيونيني **Marchionini (1995)** التصفح إلى ثلاث فئات؛ تصفح نسقي، وتصفح استكشافي، وتصفح عرضي.

ويتم التصفح النسقي أو الموجه عندما يكون المستفيد على دراية بما يبحث عنه بوضوح، كصفحة عنكبوتية بعينها، في أحد مواقع العنكبوتية العالمية، أو مصطلح محدد في معجم مصطلحات على الخط المباشر. أما المستفيدون الذين يمارسون التصفح شبه الموجه أو الاستكشافي، فأهدافهم بالنسبة للاسترجاع أقل تحديداً، أو أقل وضوحاً؛ فهم يقرؤون بسرعة أو يتفحصون للعثور على ما يسعون وراءه أو يلتمسونه؛ فمن الممكن، على سبيل المثال، أن يكون المستفيد على دراية بأن هناك فكرة مهمة قد نوقشت في تقرير مطول. وحينئذ يكون على المستفيد تصفح التقرير للعثور على الجزء المناسب منه. ويحقق التصفح غير الموجه أو التصفح العرضي أدنى درجات القابلية للتنبؤ من بين الأنواع الثلاثة للتصفح؛ فبدون حاجة بعينها إلى المعلومات، يتصفح المستفيد في نظام استرجاع المعلومات. ويمكن لأحد الأمثلة أن يكون تنقل المستفيد بسرعة من مادة إخبارية إلى أخرى، أثناء تصفحه لإحدى الصحف المعتمدة على العنكبوتية العالمية، على أمل العثور على شيء مهم للاطلاع عليه.

ولدرس التصفح من منظور مختلف، استقصى كوالسكي **Kowalski (1997)** كيف يتصفح المستفيدون ما يحصلون عليه من نتائج البحث:

١. التصفح بالترتيب الطبقي **ranking**؛ فنظراً لأن نتائج البحث في بيئة الاسترجاع الرقمية، عادة ما تقدم بترتيب طبقي، بناء على خوارزميات ترتيب طبقي معينة، يمكن للمستفيدين أن يختاروا على سبيل المثال، تصفح النتائج التي تحتل طبقة مرتفعة من حيث الصلاحية.

٢. التصفح بالمكان؛ ويدل المكان هنا على الحقول التي عادة ما تحدد معالمها، كحقل العنوان أو حقل المستخلص. والعناوين والمستخلصات بعض أمثلة مناطق المحتوى الثري بالنسبة للتصفح.

٣. التصفح بالنطاق المسلط عليه الضوء؛ فبعض نظم استرجاع المعلومات تسلط الضوء على مصطلحات معينة (كمصطلحات البحث مثلاً) وسياقاتها، لمساعدة المستفيدين في العثور على ما يبحثون عنه، من أجل تحقيق الكفاءة. وتلك النطاقات التي تسلط عليها الأضواء مرشحة بشكل مناسب للتصفح.

وفضلاً عن أنواع التصفح التي تناولها كوالسكي **Kowalski**، هناك نوعان آخران للتصفح أصبحت لهما السيادة في نظم استرجاع الإنترنت؛ أولهما التصفح بالفئة والآخر التصفح بالروابط الفائقة. ويتمثل التصفح بالفئة في نظم استرجاع الإنترنت القائمة على الدليل (مثل ياهو!). ويتم تكشيف المعلومات التي تجمعها مثل هذه النظم الخاصة باسترجاع الإنترنت، وتصنف في فئات (كالحاسبات، والتعليم، والترويح، والرياضة على سبيل المثال)؛ فالمستفيد الذي ينتوي العثور على فيلم سينمائي لمشاهدته على سبيل المثال، يمكن أن يكون من الطبيعي أن يتصفح فئة الترويح في ياهو!. أما عن التصفح بالروابط الفائقة، فإن العنكبوتية العالمية ببنيانها الفائق، لا تجعل هذا الضرب من التصفح مجدياً فحسب، وإنما مغرياً إلى حد بعيد أيضاً.

والروابط الفائقة هي النقاط الارتكازية **nodes** والمؤشرات **pointers** الكامنة في النصوص الفائقة، التي تحاكي التفكير الترابطي **associative** للإنسان. فكما أشار بوش **Bush (1945)** فإن:

> العقل البشري يعمل بالربط أو التداعي **association**؛ إذ بمادة ما تحت سيطرته ينتقل بسرعة فوراً إلى المادة التالية التي يوحي بها تداعي الأفكار، وفقاً لشبكة معقدة من المسارات بخلايا الدماغ **(p. 106)**.

وقد انعكس هذا الضرب من الترابط أو التداعي، على نحو واضح وشامل، على العنكبوتية العالمية بتطبيق الروابط الفائقة. فالعنكبوتية العالمية برمتها، تتكون من معلومات نصية، وكذلك معلومات متعددة الوسائط، متشابكة بعدد لا حصر له من الروابط الفائقة. وتلك الروابط بمثابة أدلة بالنسبة للمستفيدين من العنكبوتية العالمية، للتصفح، والملاحة، والعثور على المعلومات الرقمية **(Chu, 1997)**. ومن شأن نظم استرجاع المعلومات فائقة البناء، أن تصبح واحدة من أفضل البيئات، إن لم تكن الأفضل بالنسبة للتصفح.

استراتيجيات التصفح:

يمكن للتصفح، كالبحث، أن يتبع استراتيجيات مختلفة. وقد ذكر مارشيونيني **Marchionini** أن هناك أربع استراتيجيات للتصفح، وهي التفحص **scan**، والملاحظة **observe**، والملاحة **navigate**، والرصد **monitor**. ويمكن لاختيار استراتيجية التصفح المناسبة أن يساعد المرء على تحقيق أهداف التصفح على نحو أكثر فاعلية.

وكما ذهب مارشيونيني **Marchionini (1995)**، فإن التفحص استراتيجية تصفح من أجل التماس الأشياء محددة المعالم جيداً، في بيئة استرجاع معلومات فائقة التنظيم؛ يكون المستفيد على دراية بما يبحث عنه، وبإمكانه تفحص المعلومات طولياً أو انتقائياً. ولكي يتصفح طولياً فإنه يتصفح في فضاء المعلومات تتابعياً، وكأنه يمر على قائمة النتائج المسترجعة واحدة تلو أخرى. ويعني التفحص الانتقائي النظر فقط في قطاعات معينة من المعلومات (كالمداخل أو العناوين، والروابط الفائقة، والصور، والمحتويات ذات الألوان المختلفة في موقع العنكبوتية، على سبيل المثال) وذلك من بين تلك المحتويات التي يعرضها الموقع. ويبدو التفحص أهم استراتيجية أساس بالنسبة للتصفح النسقي.

أما الملاحظة، في مقابل ذلك، فهي الاستراتيجية الأساس التي تستخدم في التصفح الاستكشافي أو العرضي (Marchionini, 1995). وينبغي أن يكون المستفيد منتبهاً لقطاعات معينة مما تتم ملاحظته، في الوقت نفسه الذي يمكن فيه للمكونات الأخرى لبيئة استرجاع المعلومات (كالإعلانات على مواقع العنكبوتية مثلاً) أن تصرف انتباه المرء. وبعبارة أخرى، فإن مدخلات مختلفة الأنواع، يتم إرسالها إلى المستفيد من مصادر مختلفة. وعلى المستفيد أن يحتفظ بتركيزه على ما هو بصدد التماسه، ويتجاهل الشوشرة والملهيات الأخرى.

وتوازن الملاحة، كاستراتيجية تصفح، بين تأثير المستفيد والبيئة؛ إذ تقيد البيئة التصفح بتوفير مسارات التصفح المحتملة، بينما يمارس المستفيد نوعاً من التحكم باختيار أي المسارات يمكن أن يسلك (Marchionini, 1995). كذلك يمكن للملاحة، بناء على التلقيم المرتد لنظام استرجاع المعلومات، أن تستخدم على نحو أكثر من التصفح النسقي أو التصفح العرضي. وغالباً ما تستخدم الملاحظة مرتبطة بالملاحة. والرصد أو المراقبة مماثل للتفحص فيما عدا تحمله للبيئات سيئة التنظيم (Marchionini, 1995). وبينما ينظر المستفيدون في النتائج التي يسترجعها النظام، على سبيل المثال، يمكنهم أيضاً أن يولوا بعض الانتباه للتقارير الإخبارية التي يتم بثها عبر المذياع. وتتم أنشطة التماس المعلومات متوازية؛ فالرصد أو المراقبة يستخدم في الغالب الأعم مع التصفح الاستكشافي.

وقصارى القول، فإن استراتيجيات التصفح تبدو أكثر غموضاً من استراتيجيات البحث التي سبق تناولها؛ فليست هناك معايير محددة يمكن أن تساعد المستفيد في تحديد ما إذا كان التصفح ينبغي أن يتواصل أم يتوقف. وليست هناك توجيهات أو إرشادات يمكن للمستفيد اتباعها فيما يتعلق بتغيير استراتيجيات التصفح. وعوضاً عن ذلك فإنه يتعين على المستفيد النظر في عدد من العوامل الكيفية (كالرضا أو الاكتفاء والجهد المعرفي على سبيل المثال) لأغراض اتخاذ القرار. كما أنه ليس هناك أيضاً خط واضح فيما يتعلق بمتى يختار أي استراتيجية، وفي ظل أي

ظروف. فهل يمكن للتصفح كنهج للاسترجاع، أن يعمل على نحو أفضل عندما يستخدم مرتبطاً بالبحث، بدلا من استخدامه منفرداً؟ ويحاول القسم التالي الإجابة عن هذا السؤال.

تكامل البحث والتصفح في الاسترجاع:

البحث والتصفح نهجان متميزان للاسترجاع؛ فمنذ حوالي نصف قرن تقريباً، قسم لون **Luhn (1958)** طرق الاسترجاع إلى:

١. استرجاع المعلومات بالنظر في سلسلة مرتبة من التسجيلات المختزنة.

٢. استرجاع المعلومات بالبحث في سلسلة غير مرتبة من التسجيلات المختزنة.

٣. الجمع بين الطريقتين الأولى والثانية.

ولقد تغيرت مصطلحات وبنية مراصد بيانات النظم، ولا شك عبر السنين، إلا أنه لا يزال من الممكن أن نرى أن الفئة الأولى التي تناولها لون تمثل التصفح، بينما تدل الثانية على البحث. أما الضرب الثالث فهو الموضوع الذي نحن بصدد استكشافه، وهو تكامل التصفح والبحث.

المقارنة بين نهجي الاسترجاع:

لقد تناولنا بالوصف خصائص كل من التصفح والبحث في الأقسام السابقة. وكما ذهب كوكس **Cox (1992)**، فإنه من الممكن النظر إلى التصفح بوصفه الانتقال من أين إلى ماذا. والفكرة وما فيها أن المستفيد يعرف مكانه في مرصد البيانات، ويود

أن يعرف ما هو متاح هناك. وفي مقابل ذلك، يمكن النظر إلى البحث بوصفه الانتقال من ماذا إلى أين. والفكرة وما فيها أن المستفيد يعرف ما يرغب فيه، ويود معرفة أين يتاح في مرصد البيانات. وقد صرح مارشيونيني **Marchionini (1995)** بأن البحث هو الاستراتيجية الرسمية والتحليلية، بينما التصفح هو الاستراتيجية غير الرسمية أو الإيعازية **heuristic**. وبالإضافة إلى تلك الفروق، يختلف البحث والتصفح بعضهما عن بعض في الجوانب التي نعرض لها فيما يلي.

الحاجة إلى المعلومات:

الحاجة إلى المعلومات أحد المعايير الرئيسة التي تميز بين البحث والتصفح كنهجين للاسترجاع. وفيما يتعلق بالحاجة المحددة والمعروفة إلى المعلومات، يبدو البحث هو الخيار المناسب، إذ يمكن أن يساعد المستفيد فى العثور على "الإبرة" في "كومة التبن" بكفاءة. وفي مقابل ذلك يمكن للتصفح أن يستخدم لإنجاز مهام استرجاع، لمستفيدين لديهم احتياجات معلوماتية عريضة وغير واضحة المعالم. ويمكن للمرء اتباع إجراءات التصفح (كالتفحص أو الملاحة على سبيل المثال) لمعرفة ما إذا كانت هناك أي معلومات في نظام استرجاع المعلومات، مناسبة لاحتياجاته المعلوماتية.

الكفاءة واحتمالات التحسن:

ينبغي وضع الكفاءة في الاسترجاع، واحتمالات التحسن، في الحسبان عند مقارنة البحث بالتصفح. ويمكن القول بوجه عام، أن البحث سريع، ومركز، ويدخل فى الصميم، بينما التصفح يمكن أن يكون مستنفذاً للوقت، ولا يقصد به أن يكون نشاطاً مركزاً. كذلك يبدو أن المستفيد يمكن أن يكون هناك ما يصرف انتباهه في التصفح. وبينما تتاح للمستفيد أساليب متعددة لتوسعة البحث أو تضييقه، لا تتوافر

للتصفح سوى إمكانيات ضئيلة للمساعدة على الارتفاع بمستوى الأداء. وفضلاً عن ذلك فإنه لا يمكن للمستفيد الحصول على المعلومات، إلا من تلك القطاعات التي يتصفحها في النظام. ومن الممكن نظرياً للتصفح أن يتواصل، مالم يضع المستفيد نهاية له. وفي الوقت نفسه، يمكن لدقة التصفح أن تتراجع في مثل هذه العملية المطولة **(Marchionini, 1995)**.

العبء المعرفي:

من الممكن تقسيم عملية البحث برمتها، إلى حوالي ثلاث مراحل؛ التعبير عن الاستفسار، وإجراء عملية البحث، ثم تقييم النتائج. ومن الممكن للعبء المعرفي الملقي على عاتق المستفيد، أن يكون عالياً في المرحلتين الأولى والثالثة، ومنخفضاً نسبياً إذا لم يحاول المستفيد أن يكون متفاعلاً أثناء البحث. أما التصفح، في مقابل ذلك، فعملية تفاعلية بدرجة كبيرة، في جميع المراحل، بين النظام والمستفيد. ويصبح التصفح لا مبرر له إذا لم يتفاعل المستفيد مع النظام، ويظل منتبهاً لما يقدمه النظام. ومن المرهق بالنسبة للمرء أن يظل مركزاً في نهج التصفح لمدة طويلة، نظراً لأنه يتعين عليه، على نحو متواصل وسريع، تقييم المعلومات التي يتصفحها وفقاً لمحددات معينة. إلا أن التصفح يعتمد على التحقق من المعلومات ذات الصلة، لا على الاستدعاء من الذاكرة **(Large, Tedd & Hartley, 1999)**، الأمر الذي يمكن أن يخفف من العبء المعرفي الملقي على عاتق المستفيد.

الصدفة:

تكاد احتمالات الصدفة في البحث تكون منعدمة تقريباً، نظراً لأن النظام يقوم في الأساس على مضاهاة الاستفسار مقابل ما يشتمل عليه مرصد البيانات. فليس من المجدي ولا من المحتمل، بالنسبة للمستفيد أن يتفحص النظام كاملاً لاستكشاف

معلومات إضافية. وفي مقابل ذلك فإن التصفح يفسح المجال للصدفة. وقد لا يعجب المستفيد حين يجد شيئاً مفيداً، على نحو غير متوقع أثناء التصفح. إلا أن هـذه الميـزة في التصفح يمكن أن تصبح أيضاً وبسهولة، مصدراً لصرف الانتبـاه أو فقـدان الاتجـاه، وهذا قصور أشار إليه الباحثون، ومن بينهم مارشيونيني **Marchionini (1995)**.

الجـهـود:

البحث جهد منضبط، يتطلب من المرء التدريب والتمرس، حتى يكـون قـادراً على إنجازه. وفي مقابل ذلك، فإن التصفح نهج استرجاع فطـري بالنـسبة للمـستفيد، الذي لا يحتاج لإنفاق وقت في تعلمه. وفضلاً عن ذلك، لا تنطـوي عمليـة التـصفح على تعبير عن الاستفسارات، الأمر الذي يمكن أن يخفف عن كاهل المـستفيد عـبء تلك المهمة بالغة التعقد، ويكفل له التركيز ببساطة على التصفح.

ويلخـص الجـدول رقـم ٦/ ١ المقارنـة بـين البحـث والتـصفح، في خمـسة جوانب مختلفة.

الجدول رقم ٦/ ١ المقارنة بين البحث والتصفح

الجانب / النهج	الحاجة إلى المعلومات	الكفاءة	العبء المعرفي	الصدفة	الجهود
البحث	محددة ومعروفة	مرتفعة	خفيف	أقل	أكثر
التصفح	عريضة وغير مؤكدة	منخفضة	ثقيل	أكثر	أقل

النهج المتكامل:

يتبين من المقارنة، في الجدول رقم ٦/ ١، أن لكل من البحث والتصفح مظاهر قوته وأوجه قصوره. ويعمل كل نهج للاسترجاع، على نحو مناسب، في ظل شروط وظروف معينة. وعلى الرغم من أن هناك مناسبات تبدو فيها إحدى طرق الاسترجاع أكثر ملاءمة من غيرها، فإنه يمكن للتكامل بين النهجين أن يفضي إلى أداء استرجاع أفضل بوجه عام. وقد لا تكون هناك حاجة إلى البحث في كل مهمة من مهام الاسترجاع، إلا أن التصفح دائماً ما يكون ضرورياً في الحكم على صلاحية النتائج المسترجعة.

يضاف إلى ذلك أن نظم استرجاع المعلومات مصممة لتشجيع النهج المتكامل في أوساط استرجاع المعلومات. وفي نظم استرجاع المعلومات التي تطورت في وقت مبكر (ديالوج **Dialog** على سبيل المثال)، تتاح قوائم الاختيار وخيارات البحث تزامنياً، بحيث يمكن للمستفيد أن يفاضل بين التصفح والبحث وفقاً للحاجة. أما بالنسبة لنظم استرجاع الإنترنت، التي تطورت في نهاية تسعينيات القرن العشرين، فإن القاعدة هي أن يكون هناك دليل، وتقنية بحث جنباً إلى جنب. وعلى خلاف الزمن الذي ظهرت فيه نظم استرجاع الإنترنت لأول مرة للجمهور، فإن هذه النظم إما أن توفر دليلاً (مثل ياهو!) أو تقنية بحث (مثل ألتافيستا **Altavista**) وليس كليهما. ولحسن الحظ فإن كثيراً من نظم استرجاع الإنترنت، سرعان ما أدركت مزايا دعم كل من التصفح والبحث في نطاق نظام الاسترجاع الواحد، ومن ثم غيرت من ممارساتها القديمة. ولا نكاد اليوم نجد أيا من نظم استرجاع معلومات الإنترنت، يوفر خيار استرجاع واحد فقط للمستفيد.

ويلجأ المستفيدون للنهج المتكامل، نظراً لأنهم لا تتوافر لهم المقومات المناسبة لمهمة الاسترجاع المناسب فحسب، وإنما يحصلون أيضاً على المزيد من النظم نفسها. فياهو!، على سبيل المثال، يدعم البحث في نطاق فئاته القابلة للتصفح (كالفنون **Arts**

على سبيل المثال). ومن ثم، فإن البحث في نطاق فئة قابلة للتصفح، يشبه البحث في مرصد البيانات المتخصص في أحد المجالات التي تمثلها الفئة القابلة للتصفح. ويقصد بذلك القول أن البحث في ياهو!، في نطاق فئته الخاصة بالفنون، يشبه عمليات البحث في مراصد بيانات الفنون والآداب. ويصبح التصفح والبحث نشاطين متكاملين في هذه البيئة. وعلى النحو نفسه، فإن نتائج البحث في نظام لاسترجاع المعلومات (مثل كلاستي(*) **Clusty**) تتجمع تلقائياً لتيسير التصفح.

ويمكن لواحد زائد واحد أن يكون أكثر من اثنين، إذا ما تم تكوين المجموعة بحكمة. ويصدق ذلك بالنسبة لنهج الاسترجاع المتكامل الذي نتناوله هنا.

(*) أحد محركات البحث الفائقة، أو ما وراء محركات البحث **meta-search engine**. (المترجم)

المراجع

Baeza-Yates, Ricardo, and Ribeiro-Neto, Berthier. (1999). *Modern information retrieval.* New York: ACM Book Press.

Bush, Vannevar. (1945). As we may think. *Atlantic Monthly,* 176(1), 101-108.

Chu, Heting. (1997). Hyperlinks: How well do they represent the intellectual content of digital collections? *Proceedings of the 60th Annual Meeting of the American Society for Information Science,* 34, 361-369.

Cox, K. (1992). Information retrieval by browsing. In Ching-Chih Chen (Ed.), *NIT'92: Proceedings of the Fifth International Conference on New Information Technology* (pp. 69-79). West Newton, MA: MicroUse Information.

Fenichel, Carol H., and Hogan, Thomas H. (1981). *Online searching: A primer.* Marlton, NJ: Learned Information.

Herner, S. (1970). Browsing. In Allen Kent, Harold Lancour, and William Z. Nasri (Eds.), *Encyclopedia of Library and Information Science* (Vol. 3, pp. 408-415). New York: Marcel Dekker.

Hjørland, Birger. (1997). *Information seeking and subject representation.* London: Greenwood Press.

Jansen, Bernard J., Spink, Amanda, and Saracevic, Tefko. (2000). Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing & Management,* 36(2), 207-227.

Koll, Matthew. (2000). Track 3: Information Retrieval. *Bulletin of the American Society for Information Science,* 26(2), 16-18.

Korfhage, Robert R. (1997). *Information storage and retrieval.* New York: John Wiley & Sons.

Kowalski, Gerald. (1997). *Information retrieval systems: Theory and implementation.* Boston: Kluwer Academic Publishers.

Large, Andrew, Tedd, Lucy A., and Hartley, R. J. (1999). *Information seeking in the online age: Principles and practice.* London: Bowker-Saur.

Luhn, H. P. (1958). Review of information retrieval methods. Report No. RC-59. Yorktown Heights, NY: IBM Research Center.

Luhn, H. P. (1961). Selective dissemination of new scientific information with the aid of electronic processing equipment. *American Documentation,* 12(2), 131-138.

Marchionini, Gary. (1995). *Information seeking in electronic environments.* New York: Cambridge University Press. (Cambridge Series on Human-Computer Interaction 9)

Siegfried, Susan, Bates, Marcia, and Wilde, Deborah. (1993). A profile of end user searching behavior by humanities scholars: The Getty online searching project report no. 2. *Journal of the American Society for Information Science,* 44(5), 273-291.

Stielow, Frederick, and Tibbo, Helen. (Spring 1988). The negative search, online reference, and the humanities: A critical essay in library literature. *RQ,* 27(3), 358-365.

Wallace, Patricia. (1993). How do patrons search the online catalog when no one's looking? Transaction log analysis and implications for bibliographic instruction and system design. *RQ,* 33(3), 239-252.

الفصل السابع

نماذج استرجاع المعلومات

يعرَّف الأنموذج **model** بأنه "وصف تقريبي مؤقت لنظرية أو نظام، يفسر جميع خصائصه المعروفة" (**Soukhanov, et al., 1984**). وقد تطورت نماذج مختلفة في استرجاع المعلومات، على مر السنين. ويقدم هذا الفصل ويستكشف بعض النماذج المألوفة لاسترجاع المعلومات، من أجل تيسير فهم ممارسات استرجاع المعلومات وتقدير أسسها.

وهناك مستويات متعددة، يمكن فيها نمذجة استرجاع المعلومات. وتستدعى نظريات وأفكار من المجالات التخصصية الأخرى (كالمنطق البوليني، والفضاء المتجهي **vector space**، والاحتمالات على سبيل المثال) لتشكيل أسس نمذجة أنشطة استرجاع المعلومات. وقد اقترحت عدة خطط مختلفة لتصنيف جميع نماذج استرجاع المعلومات، التي تم تطويرها حتى الآن (**Baeza-Yates & Ribeiro-Neto, 1999; Belkin & Croft, 1987; Spärck Jones & Willett, 1997** على سبيل المثال). وقد توسع إنجورسن ويرفلين **Ingwersen and Järvelin (2005)** في التقسيم الهرمي لنماذج استرجاع المعلومات، الذي طوره بلكن وكروفت **Belkin and Croft (1987)**، إلى كل من المضاهاة الدقيقة **exact matching** (انظر الشكل ٤/ ٢أ) وأفضل مضاهاة **best matching** (انظر الشكل رقم ٤/ ٢ب) في كتابهما). ويركز هذا الفصل على النماذج الموجهة للنظم، أو التي تركز على النظم **system oriented**، كالمنطق البوليني، والفضاء المتجهي، والاحتمالات.

وسوف نتناول النماذج الأخرى لاسترجاع المعلومات (كالأنموذج المعرفي الموجه للمستفيد، أو الذي يركز على المستفيد **user-oriented** على سبيل المثال) في فصول أخرى، حيثما يكون ذلك مناسبا على نحو أكثر التزاما بالمنطق فيما يتعلق بالمحتوى الموضوعي.

أساس جميع نماذج استرجاع المعلومات: المضاهاة:

المضاهاة ليست من نماذج استرجاع المعلومات، وإنما تضع الأساس بالنسبة لأنشطة استرجاع المعلومات. وكما بينا في القسم ١/ ٤/ ١، فإن المضاهاة هي التقنية الأساس في استرجاع المعلومات؛ فمن الممكن إجراء المضاهاة بين المصطلحات، أو بين مقاييس التشابه كالمسافة وتردد المصطلحات. ويتم إجراء مضاهاة المصطلحات مباشرة مقابل المصطلحات المستقاة من الاستفسارات أو الوثائق أو بدائلها، أو المعينة للاستفسارات أو الوثائق أو بدائلها، بينما يتم إجراء مضاهاة مقاييس التشابه على نحو غير مباشر، مقابل المقاييس التي يتم الحصول عليها بحساب المسافة بين المتجهات، كما في أنموذج الفضاء المتجهي على سبيل المثال. ويركز القسمان الفرعيان التاليان على مناقشة هذين الضربين من المضاهاة.

مضاهاة المصطلحات:

كما بينا في القسم ٢/ ١/ ١، فإن المصطلحات التي تستعمل في تنظيم المعلومات واسترجاعها، يمكن أن تكون كلمات مفتاحية، أو واصفات، أو معرفات أو محددات للهوية. كذلك تشمل المصطلحات الكلمات أو العبارات، أو الأنواع الأخرى من وسائل التعبير. وفضلا عن ذلك يمكن مضاهاة المصطلحات بإحدى الطرق الأربع التالية، وهي المضاهاة الدقيقة، والمضاهاة الجزئية، والمضاهاة الموضعية، ومضاهاة المدى.

وتعني المضاهاة الدقيقة أن يضاهي التعبير عن الاستفسار، التعبير عن الوثيقة تمام المضاهاة في نظام استرجاع المعلومات. والبحث بمراعاة شكل الحرف الأول، والبحث بالعبارة، من أمثلة المضاهاة الدقيقة؛ "فتنقية العنكبوتية" **web filtering**، على سبيل المثال، مصطلح ورد في استفسار ما، كذلك ترد العبارة نفسها في النظام المزمع البحث فيه، حينئذ يتم الحصول على المضاهاة الدقيقة نتيجة للبحث.

وعلى عكس المضاهاة الدقيقة، فإنه في المضاهاة الجزئية، يضاهي جزء فقط من مصطلح الاستفسار مقابل طريقة التعبير عن الوثيقة في نظام استرجاع المعلومات. والبتر في البحث مثال تقليدي للمضاهاة الجزئية؛ فمن الممكن على سبيل المثال، لبحث استفسار **information technology*** (حيث النجمة * هي رمز البتر) أن يسترجع الوثائق التي تشتمل على **information technology**، و **information technologies**، و **information technologists**، نتيجة المضاهاة الجزئية.

أما المضاهاة الموضعية **positional** فتتم بوضع المعلومات الموضعية في الحسبان، في عملية المضاهاة. والبحث بالتقارب إحدى الحالات الجديرة بالنظر؛ فإذا كان استفسار البحث هو **used 1WITH store**، فإنه يمكن للنتائج المسترجعة أن تشمل الوثائق التي تشتمل على عبارات مثل "متجر الكتب المستعملة" **used book store**، و"متجر الملابس المستعملة" **used clothing store**، و"متجر الأثاث المستعمل" **used furniture store**. فالمضاهاة بين الاستفسار وطريقة التعبير عن الوثائق هنا، تتم بالكلمتين الأولى والأخيرة فقط، بينما يتم تجاهل الكلمات الواردة في موضع الوسط في عملية المضاهاة.

ومن الممكن تطبيق مضاهاة المدى **range match** في وسائل التعبير العددية **numeric** (كحجم المبيعات مثلا)، أو وسائل التعبير بالترتيب الطبيعي (مثل يناير، فبراير، ... ديسمبر). فما تتم مضاهاته في مضاهاة المدى، هو الحد الأعلى للمدى

(مطبوعات ما قبل عام ٢٠٠٢)، أو الحد الأدنى للمدى (مثل مطبوعات ما بعد عام ١٩٩٢) أو كليهما (مثل مطبوعات ما بين ١٩٩٣ و٢٠٠١). ومراصد البيانات العددية **numeric**، وتواريخ النشر أمثلة تقليدية للبحث بالمدى.

وتتعلق هذه الأنواع الأربعة المختلفة لمضاهاة المصطلحات، بالاستفسارات الأصلية، وطرق التعبير عن الوثائق، دون أي حسابات أو تحويل. وينظر إلى مضاهاة المصطلحات في الغالب الأعم، في أنموذج المنطق البوليني. أما في النماذج الأخرى لاسترجاع المعلومات (مثل الفضاء المتجهي والاحتمالات) فلا تتم مضاهاة المصطلحات الواردة في الاستفسارات، مقابل وسائل التعبير عن الوثائق، على نحو مباشر، وإنما يمكن تحويل المصطلحات إلى بعض مقاييس التشابه قبل أن تتم مضاهاتها.

مضاهاة مقاييس التشابه:

من الممكن لمضاهاة مقاييس التشابه أن تطبق بعدد من الطرق المختلفة؛ ففي أنموذج الفضاء المتجهي، على سبيل المثال، تستند المضاهاة على المسافة بين المتجهات أو درجة زاوية المتجه؛ وكلما انخفضت زاوية المتجه، ارتفعت درجة التشابه بين الاستفسارات والوثائق. أما في الأنموذج الاحتمالي فإنه من الممكن حساب التشابه اعتمادا على تردد المصطلحات، لتحديد احتمالات العلاقة الوثيقة بين الاستفسارات والوثائق.

وعند إجراء هذا النوع من المضاهاة، يتعين اختيار مقياس للتشابه خلاف المصطلحات نفسها. فبناء على هذه المقاييس تتم المضاهاة في النهاية. ومن ناحية أخرى، تكفل مضاهاة مقاييس التشابه طرقا أخرى إضافية لإنجاز مهمة الاسترجاع. ويمكن لمثل هذه الممارسات، من ناحية أخرى، أن تؤدي إلى وقوع أخطاء وحدوث شوشرة، وخصوصا في عملية الحصول على مقاييس التشابه.

وبصرف النظر عن نوعية المضاهاة، فإن المضاهاة هي الأسلوب الأساس في استرجاع المعلومات. وتوضح نماذج استرجاع المعلومات، التي تناولناها هنا كيف تتم المضاهاة في مختلف الظروف، جنبا إلى جنب مع خصائص النماذج ومزاياها وأوجه القصور التي تكتنفها.

أنموذج المنطق البوليني:

سمي أنموذج المنطق البوليني باسم جورج بول **George Boole**، الذي اقترح المنطق البوليني في منتصف القرن التاسع عشر. ويشمل المنطق البوليني ثلاث عمليات منطقية؛ هي الناتج المنطقي (×) **logical product**، والمجموع المنطقي **logical sum** (+)، والفارق المنطقي (-) **logical difference**. وتستعمل ثلاثة معاملات مقابلة، هي "و" **AND**، و"أو" **OR**، و"فيما عدا" **NOT**، للتعبير عن العمليات المنطقية في استرجاع المعلومات. وفي الطبعات المبكرة لبعض نظم استرجاع الإنترنت، كانت علامة زائد (+) تستعمل للتعبير عن المعامل "و"، الأمر الذي ربما يكون قد تسبب في بعض الخلط للمستفيدين، نظرا لأن هذه العلامة تقوم فعلا مقام المعامل "أو" في المنطق البوليني.

ويجمع الناتج المنطقي أو المعامل "و" بين مصطلحين أو أكثر في صيغة البحث، ويشترط وجود جميع المصطلحات في الوثائق المؤهلة للاسترجاع. أما المجموع المنطقي أو المعامل "أو" فيربط بين مترادفين أو أكثر، أو مصطلحات متصلة في صيغة البحث. ويمكن النظر إلى الوثائق التي تشتمل على أي من المصطلحات الواردة في صيغة البحث، بوصفها ضربات صائبة أو تسديدات **hits** أو نتائج متوقعة. أما الفارق المنطقي، أو المعامل "فيما عدا **NOT**" فيحد من عمليات البحث، باستبعاد المصطلحات التي ترد بعد المعامل "فيما عدا" في الاستفسارات. وقد وردت أمثلة بعينها لاستعمال المعاملات البولينية في القسم ٥/ ١/ ١، عندما تعرضنا لتقنية البحث البوليني.

وكما بينا في القسم ١/ ١/ ٢/ ١ فإن مورتيمر تاوبه **Motimer Taube** أدخل المنطق البوليني إلى استرجاع المعلومات. ومع التطور السريع للنظم الإلكترونية لاسترجاع المعلومات، تزايد تبني واستعمال المنطق البوليني. وعندما انتقلنا إلى العصر الرقمي، فإننا قلما نجد من نظم استرجاع المعلومات ما لا يدعم الأنموذج البوليني للبحث. وواقع الأمر أن أنموذج المنطق البوليني قد أصبح يطبق على نطاق واسع كتقنية لاسترجاع المعلومات (**Spärck Jones & Willett, 1997**). إلا أن ذلك لا يعني أن المنطق البوليني أنموذج لاسترجاع المعلومات لا يكتنفه قصور. بل العكس، إذ يمكن أن نجد مناقشات حول مظاهر القوة ومواطن الضعف في أنموذج المنطق البوليني، في كثير من الوثائق (**Chowdhury, 1999; Cooper, 1988; Frants, et al., 1999; Korfhage, 1997; Spärck Jones & Wilett, 1997** على سبيل المثال).

مظاهر قوة أنموذج المنطق البوليني:

يدل الاستخدام المكثف لأنموذج المنطق البوليني في استرجاع المعلومات، على ما لهذا الأنموذج من أهمية؛ فهو أولا يدعم المعالجة البارعة **manipulation** لمختلف الأوجه الناتجة عن تحليل الاستفسارات أو الوثائق. فبإمكان المعامل "و" **AND** الجمع بين وجهين بسيطين لتشكيل وجه مركب، وبذلك يضيق من مجال البحث. ويكفل المعامل "أو" **OR** تحديد الأوجه البديلة للاستفسار أو الوثيقة، ومن ثم توسعة مجال عملية البحث. أما المعامل "فيما عدا" **NOT** فبإمكانه عزل الأوجه المركبة في أوجه مفردة بسيطة، ومن ثم استبعاد الأوجه غير المرغوب فيها، بحيث لا تظهر في نتائج البحث النهائية. ويمكن لمثل هذه المعالجة البارعة، إذا ما طبقت على النحو السليم، أن تضفي المرونة والفاعلية على استرجاع المعلومات، على مستوى لا يمكن لأي نماذج أخرى لاسترجاع المعلومات أن تباريه.

ثانيا، تبدو نظم استرجاع المعلومات المعتمدة على المنطق البوليني فعالة من حيث التكلفة **cost-effective**، ولا غنى عنها أيضا بالنسبة للمستفيد **(Frants, et al., 1999)**. وهناك الآلاف من نظم استرجاع المعلومات القائمة على المنطق البوليني، العاملة فعلا. وبإمكان المستفيد النهائي استعمال المعاملات البولينية لتوسعة مدى عملية البحث أو تضييقها، أو استبعاد نتائج معينة من البحث. وعلى هذا النحو يبدو أنموذج المنطق البوليني مستقرا، وإن كان بعض الباحثين **(Belkin & Kroft, 1987** على سبيل المثال) يرون أن الأنموذج قد أصبح مستقرا من خلال الممارسة أكثر منه من خلال النظرية.

ثالثا، يبدو أنموذج المنطق البوليني مفهوما على نحو مناسب **(Spärck Jones & Willett, 1997)**، وإن كانت المناقشات حول ما يمكن للأنموذج أن يحققه أقل بكثير من المناقشات حول ما يعجز الأنموذج عن تحقيقه. وهذا أمر يمكن أن يكون ناتجا عن عاملين؛ أولهما أن أنموذج المنطق البوليني هو الأقدم من بين كل نماذج استرجاع المعلومات. ومن المسلم به أن مظاهر قوته تحظى بالتقدير المناسب إلى حد لا يدعو للمزيد من التحسن. والعامل الثاني هو أن المنطق البوليني، بوصفه أقدم نماذج استرجاع المعلومات، يمكن أن يُنتقد كلما ظهر أنموذج جديد لاسترجاع المعلومات. ومن الطبيعي للمرء أن يعلن أن أنموذجا جديدا لاسترجاع المعلومات، قد صمم للتغلب على أوجه قصور الأنموذج القديم. ويمكن لمصمم النظام، شأنه شأن المستفيد، أن يفضل العمل بأنموذج مفهوم بما فيه الكفاية.

رابعا، أن نظم استرجاع المعلومات القائمة على المنطق البوليني، سهلة نسبيا في بنائها، نظرا لأن الخوارزميات التي ينطوي عليها التصميم، تبدو أكثر بساطة في تطبيقها من تلك التي تقوم على نماذج أخرى لاسترجاع المعلومات. ويمكن لذلك أن يسهم جزئيا أيضا في اتساع نطاق تطبيق النظم البولينية لاسترجاع المعلومات. ونظرا

لمظاهر القوة التي سبق ذكرها (كأسلوب المعالجة البارعة، والاستخدام على نطاق واسع على سبيل المثال) فإن أنموذج المنطق البوليني دائما ما يحظى بقدر كبير من الاهتمام حيثما تكون نماذج استرجاع المعلومات موضوعا للتدقيق التحليلي.

أوجه قصور أنموذج المنطق البوليني:

حظيت أوجه قصور أنموذج المنطق البوليني، كما سبق أن بينا، بالدرس والاستكشاف على نحو ضافٍ في كثير من الوثائق (**Chowdhury, 1999; Cooper, 1988; Frants, et al., 1999; Korfhage, 1997; Spärck Jones & Willett, 1997** على سبيل المثال) . والمناقشة التالية هي خلاصة أوجه قصور ذلك الأنموذج.

أولا، من الصعب على المستفيد إجراء عمليات البحث البوليني بدون درجة مناسبة من التمرين والتمرس. وتكمن الصعوبة في جانبين: (١) من الصعب بالنسبة للمستفيد اختيار المعامل البوليني المناسب؛ فدائما ما يكون هناك خلط بشأن المعاملين "و" **AND** و"أو" **OR**، في أوساط المستفيدين، نظرا لأن لهاتين الكلمتين معانٍ مختلفة في المغزى التقليدي أو الدلالة المشتركة؛ فكلمة **and** تعني تقليديا زائد **plus**، كما في "إنهم سيبحثون عن هذا الموضوع في ألتافيستا **AltaVista**" و"جوجل **Google**"، على سبيل المثال، بينما تعني كلمة **or** أيًا من **either one**، كما في "سيبحثون عن هذا الموضوع في ألتافيستا" أو "جوجل"، على سبيل المثال. ومن الممكن نظريا البحث في مواقع في الحالة الأولى، أكثر مما يحدث في الحالة الأخيرة. وعادة ما يفكر بعض المستفيدين على النحو نفسه، عندما يصوغون الاستفسارات البولينية، أي أنهم يستعملون المعامل **AND** عندما يكونوا راغبين في الحصول على نتائج كثيرة، ويستعملون المعامل **Or** عندما يكونوا راغبين في نتائج قليلة. ومما لا شك فيه أن المنطق البوليني لا يعمل على هذا النحو. ومن المحتمل لهذا الفارق أن يؤدي إلى

اختيار المستفيد للمعاملات غير المناسبة، أي يستعمل المعاملين AND وOR في السياقات غير المناسبة. (٢) من الصعب بالنسبة للمستفيد أن يتبع تسلسل المعالجة على نحو صحيح، في عمليات البحث البوليني المركبة. وكما أوضحنا في القسم ٥/ ١/ ١/ ١ فإن عمليات البحث البوليني المركبة تنطوي على أكثر من ضرب واحد من المعاملات البولينية، ويتم تحديد التسلسل الطبيعي للبحث على النحو التالي: يأتي المعامل "فيما عدا" NOT أولا، ويأتي المعامل "و" AND ثانيا، بينما يأتي المعامل "أو" OR أخيرا. إلا أنه من الممكن تغيير التسلسل الطبيعي للمعالجة، بالنسبة لعمليات البحث البوليني المركبة، باستعمال الأقواس حينما تدعو الحاجة. غير أن مثل هذا التغيير معقد ويتطلب الكثير، وخصوصا بالنسبة للمبتدئين. فقد لا يكون المستفيد قادرا على التكيف مع القواعد الاصطناعية (كمعالجة الأقواس الطولية أو الموغلة innermost قبل غيرها، على سبيل المثال) من أجل تغيير التسلسل الطبيعي للمعالجة، بالنسبة لعمليات البحث البوليني المركبة، ما لم تتح له فرصة التدريب والتمرس الملائمة.

ثانيا، من الصعب التعبير عن العلاقات غير البولينية بين المصطلحات (علاقة السببية على سبيل المثال) نظرا لأن هذا الأسلوب غير متوافر ببساطة في الأنموذج. لنفترض أن مستفيدا يود العثور على بعض المعلومات حول تطبيقات الحاسب في التعليم. حينئذ تتم صياغة استفسار بحث باستعمال المعامل البوليني "و" AND: "الحاسبات والتعليم" computers AND education. ولا يرد مصطلح "التطبيقات" application في الاستفسار، نظرا لأن هذه العلاقة من المفترض أن يتم التعبير عنها عن طريق المعاملات البولينية، إلا أن أيا من هذه المعاملات لا ينهض بهذه الوظيفة. ومن ثم فإنه بإجراء عملية البحث "الحاسبات والتعليم" يمكن للمستفيد أن يحصل على معلومات، لا حول استخدام الحاسبات في التعليم فحسب، وإنما حول

تعليم الحاسبات أيضا. ويمكن ولا شك لـصيغ البحـث البـوليني أن تكـون في غايـة التعقد. إلا أنه لا يزال هناك ثلاثة معاملات فقط متاحة للمستفيد، لكي يربط بها بـين المفاهيم التي تتضمنها عملية البحث. ومما لا شك فيه بوجـه عـام، أنـه كلـما ازدادت صيغة البحث تعقدا زادت احتمالات تسببها في التفسيرات غير الصحيحة، نظرا لعجز أنموذج المنطق البوليني عن التعبـير عـن علاقـات أخـرى خـلاف علاقـات المنطـق البوليني بين المصطلحات.

ثالثا، ليست هناك تقنية للوزن متاحة في أنموذج المنطق البوليني، لبيان الأهميـة النـسبية لمختلـف المفـاهيم في استفـسار البحـث. ومـن المفـترض أن تكـون لجميـع المصطلحات أو المفاهيم في الاستفسارات البولينية، الدرجة نفسها من الأهمية، الأمـر الذي لا يصدق دائما في اسـترجاع المعلومـات. فـإذا افترضـنا عـلى سـبيل المثـال، أن استفسار البحــث هــو "الوصـول إلى المعلومـات والأمـن" **information access AND security**، ويود المـستفيد أن يركـز أكثـر عـلى مـصطلح "الأمـن". وبعبـارة أخرى، يود المستفيد الحـصول عـلى معلومـات حـول قـضية الأمـن في الوصـول إلى المعلومات، بدلا من معالجة المصطلحين بوصفهما متساويين في الأهمية. إلا أنه في نظام بوليني لاسترجاع المعلومات، قد لا يكون من الممكن تلبية توقع المستفيد، نظرا لعـدم إمكان تخصيص أوزان لمصطلحات الاستفسارات.

رابعا، من المستحيل التعبير عن الصلاحية الجزئيـة لنتـائج عمليـة البحـث، في أنموذج المنطق البـوليني، نظـرا لأن هـذا الأنمـوذج ببـساطة، يقـسم جميـع البيانـات المتوافرة في النظام إلى فئتين؛ صالحة عندما تكون هناك مضاهاة بين الاستفسار وبدائل الوثائق، وغير صالحة عندما لا تكون هناك مضاهاة. وليست هناك فئة وسط، ومن ثم فإن نظم المنطق البوليني لا يمكن أن تقدم نتائج مرتبة تنازليا وفقا للصلاحية. ونظـرا لأنه ليست هناك مخرجات مرتبة طبقيا، فإنه لا يمكن للمستفيد أن يقرر، عـلى سـبيل

المثال، استرجاع الوثائق الخمسة عشر المتربعة على القمة. وعلى المستفيد غربلة جميع النتائج، التي أحيانا ما تعد بالآلاف أو أكثر، التي لا يقدمها النظام البوليني في ترتيب معين للصلاحية. وعلى هذا النحو فإنه لا سيطرة للمستفيد على حجم المخرجات في عمليات البحث البولينية.

خامسا، يمكن للمستفيد أن يحصل على مخرجات لا طائل من ورائها على الإطلاق، أو مخرجات تشكل عبئا زائدا عليه، عند إجراء البحث البوليني. ويمكن للمخرجات أن تكون لا طائل من ورائها إذا كان استفسار البحث مقيدا إلى حد بعيد، كما هو الحال على سبيل المثال، عندما تكون هناك عدة مصطلحات مرتبطة معا بالمعامل "و". ومن ناحية أخرى، يمكن للمخرجات التي تشكل عبئا زائدا أن تتأتى فعلا، إذا كان استفسار البحث عريضا، كما هو الحال على سبيل المثال، عندما ترتبط عدة مصطلحات معا بالمعامل "أو" **OR**. وبإمكان المستفيد تعديل استفسار البحث لزيادة مخرجات عمليات البحث أو الحد منها. إلا أن الناتج حينئذ لن يكون ما كان يبحث عنه المستفيد من البداية، وإنما مخرجات الاستفسار المعدل.

وللتغلب على ما يكتنف الأنموذج البوليني من قصور، اقترح كوبر **Cooper (1988)** بعض الحلول الممكنة، كصياغة الاستفسارات خالية من المعاملات البولينية، لتجنب عدم تعاطف الاستفسارات البولينية مع المستفيد. ومن المقترحات الممتازة لكوبر استمارة البحث **search form**، على الرغم من أن هذا الاقتراح لم يطبق على نطاق واسع حتى نهاية تسعينيات القرن العشرين. وفضلا عن ذلك تم أيضا تطوير عدد لا بأس به من الخوارزميات والطرق، بما في ذلك نماذج استرجاع المعلومات، من أجل توفير مخرجات مرتبة طبقيا، ومدخلات موزونة، وما شابه ذلك. إلا أن "نماذج الاسترجاع الأخرى، التي يمكن أن تكون جديرة بالاهتمام نظريا، لم تحقق في التطبيق نتائج تفوق على نحو بارز، تلك النتائج الخاصة بنظم الاسترجاع البولينية"**(Korfhage, 1997, P. 63)** .

أنموذج الفضاء المتجهي:

تطور أنموذج الفضاء المتجهي **vector space**، الذي يعرف أيضا بالتجهيز المتجهي أو استرجاع الناتج المتجهي، على يدي سالتون **Salton** وزملائه، الذين بنوا نظام معالجة النصوص واسترجاعها **System for the manipulation and retrieval of texts**(سمارت **SMART**)، ليكون بمثابة أساس لإجراء سلسلة طويلة من تجارب استرجاع المعلومات (**Salton, 1968** على سبيل المثال). كذلك صممت سلسلة من أساليب استرجاع المعلومات (وزن المصطلحات، والمخرجات المرتبة طبقيا، والتلقيم المرتد للصلاحية، على سبيل المثال) في سياق عملية بناء الأنموذج. وإلى جانب المنطق البوليني، كان لأنموذج الفضاء المتجهي أكبر تأثير على تطور استرجاع المعلومات، ونظم استرجاع المعلومات العاملة **(Spärck Jones & Willett, 1997)**.

وفي أنموذج الفضاء المتجهي، يتم تحديد كل مصطلح بوصفه أحد الأبعاد، بينما يتم التعبير عن كل استفسار أو وثيقة بوصفه متجها. ويتكون المتجه فعلا من قائمة من قيم المصطلحات تمثل مادة ما (أي استفسار أو وثيقة). ومن الممكن لقيم المصطلحات في المتجه أن تكون ثنائية أو موزونة. ويمكن للقيم الثنائية أن تكون واحد أو صفر، حيث يدل الواحد على وجود المصطلح في المادة. أما القيم الموزونة فتشمل أعدادا إيجابية حقيقية (مثل ١.٥، و٣.٠، و٢.٤، و٥.٦). وتقابل القيمة الموزونة لكل مصطلح الأهمية النسبية لذلك المصطلح، في تمثيل هذه المادة **(Kowalski, 1997)**. ويمكن للخطة التي تستخدم لوزن المصطلحات في أنموذج الفضاء المتجهي، أن تكون موضوعية (كتردد المصطلحات مثلا) أو ذاتية أي غير موضوعية (كإدراك المستفيد أو فهمه مثلا). وطرق الوزن التي عالجناها في القسم ٥/ ١/ ٢/ ٢ قابلة نظريا للتطبيق في تحديد الأوزان للمصطلحات في متجه ما. ولكل خوارزمية للوزن مزاياها

وعيوبها. وقد ناقشن كورفيج **Korfhage (1997)** الطريقة القائمة على تردد المصطلحات تفصيلا، وكذلك الطريقة القائمة على إدراك المستفيد، بالنسبة لتحديد الأوزان للمصطلحات في متجه ما.

ويتساوى عدد الأبعاد في متجه استفسار أو وثيقة ما، مع عدد المصطلحات المختلفة التي تمثل المادة. وتشكل جميع متجهات الاستفسار والوثيقة الفضاء متعدد الأبعاد. ومن ثم فإن المجموعة الكاملة لقيم المصطلحات في متجه ما، تدل على موقع الاستفسار أو الوثيقة التي تمثلها في الفضاء **(Spärck Jones & Willett, 1997)**.

ويعني إجراء بحث في نظام الفضاء المتجهي لاسترجاع المعلومات، مراجعة المسافة مبينة كزاوية، بين متجهي الاستفسار والوثيقة في الفضاء. ويوضح أنموذج الفضاء المتجهي أوجه التشابه بين الوثيقة والاستفسار (أو بين أي وثيقتين) بمقارنة أبعادهما المقابلة، وبحساب مقاييس التشابه هذه، بوصفها معامل جيب التمام **cosine coefficient**. فإذا كان كل من الاستفسار والوثيقة حول موضوع مماثل، فإن الزاوية بين متجهيهما ينبغي أن تكون صغيرة، أما إذا كان الاستفسار والوثيقة حول موضوعات مختلفة، فإن الزاوية بين متجهيهما ينبغي أن تكون كبيرة **(Lesk, 1997)**. كذلك يمكن قياس أوجه التشابه بين الوثائق بالطريقة نفسها.

مظاهر قوة أنموذج الفضاء المتجهي:

يكفل أنموذج الفضاء المتجهي، كما ذهب سبارك جونز وولت **Spärck Jones and Willett (1997)**، أساسا موحدا لمدى واسع من عمليات الاسترجاع، يشمل التكشيف، والتلقيم المرتد للصلاحية، وتصنيف الوثائق. وتتضح أوجه قوة هذا الأنموذج على أحسن وجه عندما تقارن بمظاهر قصور أنموذج المنطق البوليني.

أولا، لم يعد المستفيد بحاجة لأن يلم بالمنطق البوليني ويطبقه، عند إجراء عمليات البحث في نظام استرجاع معلومات الفضاء المتجهي. إلا أن ما يتعين على المستفيد عمله هو أن يختار ببساطة عدة مصطلحات بناء على حاجته إلى المعلومات.

ثانيا، يمكن للمصطلحات أو المفاهيم التي يقع عليها الاختيار لتمثيل الاستفسار أو الوثيقة، أن توزن لبيان أهميتها النسبية في المتجه. فإذا كان الاستفسار أو الوثيقة يهتم على سبيل المثال، أساسا بقضية الأمن في المشابكة، فإنه يمكن تخصيص وزن مرتفع لمصطلح "الأمن" **security**، ووزن منخفض لمصطلح "المشابكة" **networking**. وعلى هذا النحو لن يعامل هذان المصطلحان بوصفهما متساويين في الأهمية في المتجه، ومن ثم يمكن الخروج بتمثيل ملائم للاستفسار أو الوثيقة.

ثالثا، ترتب مخرجات عمليات بحث الفضاء المتجهي طبقيا، تبعا لتناقص الصلاحية، نظرا لأن الأنموذج يعبر عن التشابه أو العلاقة بين الوثائق، أو بين الاستفسار ووثيقة ما على مقياس (من ٠.٠ إلى ١.٠، حيث ١.٠ هو الأوثق صلة)، على سبيل المثال، عوضا عن التقسيم الثنائي (أي إما على صلة وإما ليس على صلة) كما يتم في أنموذج المنطق البوليني. ومن ثم فإنه بإمكان المستفيد أن يراجع، على سبيل المثال، الوثائق العشر المتربعة على القمة دون غيرها، وهو على يقين من أن البقية، أي النتائج التي لم يراجعها، لا يمكن أن تكون أوثق صلة بالاستفسار من الوثائق العشر التي استعرضها. وواقع الأمر أن قدرة المستفيد على الحد من حجم مخرجات الاسترجاع، ترتفع بالمقارنة بأنموذج المنطق البوليني. وتقتصد المخرجات المرتبة طبقيا، في وقت المستفيد وطاقته في عملية استرجاع المعلومات، نظرا لأنه ليس على المستفيد النظر في جميع الوثائق، التي تبلغ الآلاف في غالب الأحيان، قبل أن يقرر أي الوثائق أوثق صلة من غيرها. وبإمكان النظام اختيار وتقديم أكثر الوثائق صلاحية بالنسبة للمستفيد تلقائيا.

رابعا، يشكل التلقيم المرتد للصلاحية تقنية يـتم تطبيقهـا في أنمـوذج الفـضاء المتجهي، من أجل الارتفاع بمستوى أداء الاسترجاع. وبناء على الناتج الصالح الذي سبق استرجاعه، يمكن لنظام استرجاع المعلومات المعتمد عـلى الفـضاء المتجهـي، أن يكون قادرا على تعـديل متجـه الاستفـسار تلقائيـا، وتقـديم أكثـر النتـائج صـلاحية للمستفيد. ويمكن للتلقيم المرتد الخاص بالصلاحية أن يتم دون تدخل المستفيد، كـما يمكن تكرار العملية عدة مرات وفقا للـضرورة. ومـن الممكـن للمـستفيد أن يفعِّـل العملية في نظم استرجاع معلومات الإنترنت، وذلـك ببـساطة بالاتجـاه نحـو الخيـار الموسوم بعبارة "المزيد مما يشبه هذا **More like this**"، أو "المزيد من النتائج المماثلـة **More similar results**"، على سبيل المثال.

وكما تبين لنا في هذه المناقشة، فإن أنموذج الفضاء المتجهي قد تغلب على بعض أوجه القصور المرتبطة بأنموذج المنطق البوليني. إلا أن أنموذج الفـضاء المتجهـي قـد أدخل أيضا بعض المشكلات الجديدة في مجال استرجاع المعلومات.

أوجه قصور أنموذج الفضاء المتجهي:

أول أوجـه قـصور أنمـوذج الفـضاء المتجهـي هـو افـتراض الانفـصال بـين المصطلحات التي يتم اختيارها لوصف متجـه مـا. وعنـد النظـر في أنمـوذج المنطـق البوليني يتبين أن الأنموذج لا يمكن أن يعبر عن ضروب أخرى من العلاقات، خارج نطاق المنطق البوليني. إلا أن أنمـوذج الفـضاء المتجهـي لا يـستطيع التعبـير عـن أي علاقة، بما في ذلك العلاقات البولينية بين المصطلحات. وبالأحرى فإنه ينطـوي عـلى الافتراض بأن تكون المصطلحات التـي تـشكل الأسـاس بالنـسبة للفـضاء المتجهـي متعامدة أو مستقلة عن بعضها البعض. ويبدو أن هذا الافتراض غـير صـحيح؛ فـإذا كانـت المـصطلحات التـي اختـيرت لوصـف متجـه وثيقـة مـا، هـي "الـسيارات" و"التصدير" و"الاستيراد" **automobiles, export and import**، فكيف يمكـن

أن نفترض أنه ليست ثمة علاقة على الإطلاق بين المصطلحات الواردة؟ وقد انتقد هذا القصور بشدة في أوساط أولئك الذين يهتمون بالأنموذج.

ومظهر القصور الثاني في أنموذج الفضاء المتجهي، هو صعوبة تحديد علاقات الترادف أو العلاقات النحوية صراحة، نظرا لغياب المعاملات البولينية ومعاملات التقارب. ولكي نكون أكثر تحديدا، فإن الأنموذج لا يدعم المعاملات البولينية ومعاملات التقارب. ومن ثم فإن المعامل "أو" **OR** لا يمكن أن يستعمل للدلالة على المترادفات (كما في **cars OR automobiles**). كما أن المعامل "مع" **WITH** لا يمكن أن يستعمل لتكوين عبارات (كما في **information WITH retrieval**) في متجه ما. إلا أنه في عمليات البحث الفعلية، هناك دائما ظروف تدعو الحاجة فيها إلى المترادفات أو العبارات، لتمثيل الاستفسارات أو الوثائق. ومن ثم فإنه من الصعب على المستفيد إجراء عمليات بحث بدون الإمكانيات البولينية، ومعامل التقارب في نظم استرجاع المعلومات القائمة على الفضاء المتجهي، عندما لا يكون قد تم التعبير عن علاقة الترادف والعلاقات النحوية بين المصطلحات.

أما مظهر القصور الثالث لأنموذج الفضاء المتجهي، فيتصل بتقنيته في الوزن، التي يمكن أن تكون غير موضوعية ومعقدة؛ فعملية الوزن تكون غير موضوعية عندما يطلب من المستفيد تحديد أوزان للمصطلحات، وخصوصا مصطلحات الاستفسار، وذلك بناء على فهمه أو تقديره الخاص. ومن المفترض في المستفيد أن يقدر الأهمية النسبية التي يمكن أن يحظى بها المصطلح في عملية البحث، ثم يحدد وزنا له. ولا مناص من الذاتية في ممارسة تلك العملية. ومن ناحية أخرى، يمكن للوزن أن يكون معقدا، بحيث لا يمكن لأي خوارزمية للوزن أن تعمل بدون خلل، كما أن العثور على أفضل خوارزمية لبيئة استرجاع معينة ليس بالمهمة السهلة. يضاف إلى ذلك، أن مراصد البيانات في نظم استرجاع المعلومات تتسم بالدينامية، نظرا لحاجتها

إلى التحديث المستمر. وهكذا، يمكن لأوزان المصطلحات أن تتغير تبعا لذلك، نظرا لأن المحددات التي تستند إليها طريقة الوزن (كتردد المصطلحات مثلا) قد تغيرت. وقد لخص كوالسكي **Kowalski (1997)** عددا من طرق التعامل مع مراصد البيانات التي تتغير ديناميكيا، وما لها من تأثير على خوارزميات الوزن. إلا أنه كما أشار كوالسكي نفسه، ليس من بين هذه الطرق ما يمكن أن يطبق دون بعض المخاوف (كالتكلفة مثلا).

وإلى جانب أوجه القصور الثلاثة التي تناولناها توا، يحتاج أنموذج الفضاء المتجهي أيضا إلى عدة مصطلحات لتمثيل استفسار ما أو وثيقة ما، حتى يمكن للمتجه أن يكون قادرا على التمييز بما فيه الكفاية، بالنسبة لأداء الاسترجاع المناسب. وفي مقابل ذلك، "يمكن لمصطلحين فقط أو ثلاثة ترتبط بالمعامل "و" **AND** أن تكون كافية في بيئة بولينية، للحصول على مخرجات عالية الجودة" **(Spärck Jones & Willett, 1997, P. 259)**. ويمكن لتخصيص المزيد من المصطلحات، في أنموذج الفضاء المتجهي، أن يؤدي إلى المزيد من التكلفة. وفضلا عن ذلك، يفتقر الأنموذج إلى التبرير النظري لبعض عمليات معالجة المتجهات؛ فاختيار أسلوب بعينه لحساب تشابه المتجه، في استعمال أنموذج استرجاع المعلومات، على سبيل المثال، لا يتقرر بناء على أي اعتبارات نظرية، وإنما يترك الأمر للمستفيد **(Salton, 1989)**.

وينبغي مثاليا، أن يضع أنموذج الفضاء المتجهي، متجهات الوثائق على نحو يؤدي إلى تعنقد أو تجمع **clustered** تلك الوثائق الصالحة لاستفسارات معينة معا، بينما يمكن لتلك الوثائق غير المرغوب فيها تزامنيا على الإطلاق، أن تبدو منفصلة في الفضاء **(Salton, Wong & Yang, 1975)**. إلا أن مضاهاة استفسار ما، مقابل مجموعات أو عناقيد من الوثائق، المسماة بالفرضية العنقودية **cluster hypothesis**، لم تتسن بعد **(Spärck Jones & Willett, 1997)**.

ولم يطبق أنموذج الفضاء المتجهي في كثير من نظم استرجاع المعلومات العاملة إلا بعد ظهور نظم استرجاع الإنترنت. ويتيح نظام سمارت **SMART** الذي يعد بمثابة حقل الاختبار، بالنسبة لهذا الأنموذج، فرصة النمو والنضج. ولقد أدى ابتكار أنموذج الفضاء المتجهي إلى إثراء البحث والتطوير في مجال استرجاع المعلومات إلى حد بعيد.

أنموذج الاحتمالات:

لقد أدخل مارون وكونز **Maron and Kuhns (1960)** أنموذج الاحتمالات، الذي تطور فيما بعد على يدي روبر تصون وعدد آخر من الباحثين (**Robertson & Spärck Jones, 1976** على سبيل المثال). وكما ذهب سبارك جونز وولت **Spärck Jones and Willett (1997)**:

> إن الأساس المنطقي لإدخال المفاهيم الاحتمالية واضح؛ فنظم استرجاع المعلومات تتعامل مع اللغة الطبيعية، وهذه تفتقر إلى الدقة إلى حد بعيد، بحيث تكفل للنظام القدرة على أن يقرر أي الوثائق يمكن أن تكون صالحة، بالنسبة لاستفسار معين (وهذا موقف يتناقض بوضوح مع عمليات الاسترجاع غير الغامضة، اللازمة للبحث في نظم إدارة مراصد البيانات العددية). **(p. 259)**

ويطبق هذا الأنموذج نظرية الاحتمالات، أي الحالة التي يحتمل حدوثها من صفر بالمئة إلى ١٠٠ بالمئة (٠.٠ – ١.٠) لاسترجاع المعلومات. وهو يضع في الحسبان عنصر عدم التأكد في عملية استرجاع المعلومات، أي عدم التأكد فيما يتعلق بما إذا

كانت الوثائق المسترجعة بواسطة النظام، صالحة بالنسبة لاستفسار معين أم لا **(Bookstein, 1985)**. ويهدف هذا الأنموذج إلى تقدير وحساب احتمال أن تكون وثيقة ما صالحة لاستفسار معين، بناء على بعض الطرق. وتدل الحالة في هذا السياق الخاص باسترجاع المعلومات، على احتمالات العلاقة بين الاستفسارات والوثائق. وعلى عكس النماذج الأخرى لاسترجاع المعلومات، لا يعامل أنموذج الاحتمالات العلاقة بوصفها مقياسا دقيقا للإخفاق أو المضاهاة **miss-or-match**، وإنما يتم التعبير عن العلاقة بناء على الاحتمالات. فمن الممكن على سبيل المثال للأنموذج أن يحيط مستفيدا ما علما بأن الوثيقة "و" احتمالات ارتباطها بالاستفسار "أ" ٣٥ بالمئة.

ويتبع هذا الأنموذج طرقا مختلفة لتحديد احتمالات الارتباط بين الاستفسارات الوثائق. ويتم الحكم على الارتباط في أنموذج الاحتمالات وفقا للتشابه بين الاستفسارات والوثائق. ويتوقف حكم الارتباط أيضا على تردد المصطلحات **term frequency**. ويمكن القول بوجه عام، أنه كلما زاد التشابه القائم بين الاستفسارات والوثائق، زادت احتمالات أن تكون تلك الوثائق صالحة بالنسبة للاستفسارات. ومن المفترض في النظام الاحتمالي لاسترجاع المعلومات، في النتائج التي يتم الحصول عليها بالنسبة لأي استفسار، أن تتكون من الوثائق التي تفي بشروط ذلك الاستفسار، باحتمالات أعلى من حد معين **(Korfhage, 1997)**.

مظاهر قوة الأنموذج الاحتمالي:

بالمقارنة بأنموذج المنطق البوليني أو أنموذج الفضاء المتجهي، يتمتع أنموذج الاحتمالات بمظاهر القوة التالية:

أولا، يكفل هذا الأنموذج "دليلا مهما للتقسيم الفئوي لعمليات الاسترجاع، فضلا عن التبرير النظري للممارسات التي كانت تنبع فيما مضى، على أساس تجريبي أو

خبر **empirical**،كإدخال نظم معينة لوزن المصطلحات على سبيل المثال" **(Salton, 1989, pp. 348-349)**. وتتسم عمليات الاسترجاع بدرجة من عدم التيقن، عند الحكم على مدى قوة العلاقة بين الاستفسارات والوثائق. وفضلا عن ذلك تتقرر العمليات الجوهرية للأنموذج، كقياس التشابه بين الاستفسارات والوثائق، بواسطة الأنموذج نفسه، بدلا من أن تتقرر ببعض القرارات التعسفية، كما في أنموذج الفضاء المتجهي.

ثانيا، يتضمن الأنموذج علاقات تبعية المصطلحات وما يرتبط بها (أي أن حدوث واقعة ما من شأنه أن يؤدي إلى حدوث الأخرى في تطبيقات هذا الأنموذج). ولم يعد المستفيدون بحاجة لافتراض استقلال المصطلحات بعضهما عن بعض، كما كانوا في أنموذج الفضاء المتجهي، وهو افتراض لا يحظى بالدعم في الممارسة العملية. كذلك يزن هذا الأنموذج مصطلحات الاستفسارات والارتباط بين الاستفسارات والوثائق، وبذلك يتمكن المستفيدون من معرفة الأهمية النسبية للمصطلحات أو الارتباط، في مهمة الاسترجاع. كذلك يوفر هذا الأنموذج المخرجات المرتبة طبقيا، نظرا لأنه يفترض أن تكون المهمة الأساس لنظام استرجاع المعلومات، هي ترتيب الوثائق التي تتكون منها المجموعة، وفقا لتناقص احتمالات الارتباط بحاجة المستفيد إلى المعلومات **(Spärck Jones & Willett, 1997)**. ويسمى هذا الافتراض مبدأ الترتيب الطبقي للاحتمالات. وبناء على مبدأ الترتيب الطبقي للاحتمالات، يمكن للمستفيد أن يمارس قدرا من التحكم في حجم مخرجات الاسترجاع. ويتم التعبير عن كل من الوزن والترتيب الطبقي بالاحتمالات.

ثالثا، بإمكان هذا الأنموذج الإفادة من معلومات التلقيم المرتد، لتطوير طرق تستند إلى أساس متين **(Bookstein, 1985)**. وعلى النحو نفسه أشار كوالسكي **Kowalski (1997)** إلى أنه من الممكن لهذا الأنموذج أن يتحقق بدقة، من افتراضاته

الضعيفة، ويعمل على تقويتها. وتضفي خاصية التحسن الذاتي عنصرا إيجابيا آخر على أنموذج الاحتمالات.

رابعا، لا يستخدم أنموذج الاحتمالات في شكله الأصلي، إمكانيات المنطق البوليني، التي يمكن أن يجد المستفيدون صعوبة في تطبيقها. ومن ثم فإن نظم استرجاع المعلومات القائمة على أنموذج الاحتمالات، أكثر تعاطفا مع المستفيد، من نظم الاسترجاع القائمة على المنطق البوليني.

مظاهر قصور أنموذج الاحتمالات:

حظيت مظاهر قصور أنموذج الاحتمالات بالدرس والتدقيق أيضا، من زوايا مختلفة منذ ابتكاره. وفيما يلي نلخص هذه المظاهر بإيجاز.

أولا، على الرغم من أن قيمة الصلاحية في هذا الأنموذج متواصلة، بدلا من التقسيم الثنائي صفر أو واحد، كما في الأنموذج البوليني، فإن أنموذج الاحتمالات يفترض أن تتسم الصلاحية بالثنائية، أي أن:

$$\mathbf{Pr}_{(nonrel)} = \mathbf{1- Pr}_{(rel)}$$

حيث $\mathbf{Pr}_{(rel)}$ هو احتمال الصلاحية، و$\mathbf{Pr}_{(nonrel)}$ هو احتمال عدم الصلاحية. وبعبارة أخرى فإن القيمة بالنسبة لاحتمال عدم الصلاحية ثابتة بمجرد حساب احتمالات الصلاحية، وبالتالي يتم استبعاد عامل عدم التيقن الجوهري، من عملية استرجاع المعلومات. فكما أشار روبرتصون **Robertson (1977)**، فإن افتراض التقسيم الثنائي افتراض قوي، ومن المؤكد تقريبا أنه غير صالح بوجه عام.

ثانيا، لا يمكن لأنموذج الاحتمالات أن يؤدي إلى الارتفاع بمستوى فاعلية الاسترجاع على نحو ملحوظ؛ فالنتائج التي أمكن الحصول عليها باستعمال أنموذج الاحتمالات، وإن كانت جيدة، لم تكن أفضل بما فيه الكفاية، من تلك المسترجعة من

نظم المنطق البوليني أو نظم الفضاء المتجهي..ومن الممكن طبيعيا أن يثار تساؤل حول ضرورة وجود أنموذج إضافي لاسترجاع المعلومات، بينما لا يمكن لأدائه أن يكون مختلفا عن أداء النماذج القائمة.

وبالإضافة إلى مظهري القصور اللّذين تناولناهما هنا، يكتنف هذا الأنموذج مظاهر قصور أخرى؛ فالتمرس في الرياضيات، على سبيل المثال، لا غنى عنه للإحاطة باستعمال نظرية الاحتمالات، الأمر الذي يمكن أن يؤثر سلبا في سهولة التعامل مع الأنموذج. يضاف إلى ذلك أن هناك كثيرا من الأشكال المختلفة للأنموذج؛ فهناك بعض مظاهر الاختلاف حول أي الطرق هي الأفضل، وكيف يمكن حسم مختلف القضايا التقنية، حتى وإن كان هناك إجماع أو اتفاق حول الأنموذج العام (**Bookstein, 1985**). وعلى نحو مماثل لأنموذج الفضاء المتجهي، كان أنموذج الاحتمالات، حتى ظهور نظم استرجاع الإنترنت، يستعمل للأغراض التجريبية، أكثر مما كان يطبق في النظم العاملة لاسترجاع المعلومات.

امتدادات النماذج الرئيسة لاسترجاع المعلومات:

عادة ما ينظر لكل من المنطق البوليني، والفضاء المتجهي، والاحتمالات، بوصفها النماذج الثلاثة الرئيسة لاسترجاع المعلومات. ولكل من هذه النماذج امتداداته؛ فالمنطق البوليني الممتد، على سبيل المثال، امتداد لأنموذج المنطق البوليني وأنموذج الفضاء المتجهي مجتمعين. كذلك تستند الفئة الغامضة أو المشوشة **fuzzy set** إلى أنموذج المنطق البوليني، إلا أنها - فضلا عن ذلك- تدخل نظرية الفئات في استرجاع المعلومات. والتكشيف الدلالي الكامن **latent**، من اشتقاقات أنموذج الفضاء المتجهي؛ إذ يهدف إلى العثور على مجموعة أصغر حجما من الأبعاد والقيم، التي يمكن أن تستعمل بوصفها بديلا عن المصفوفة **matrix** الأصلية **(Lesk,**

(1997. وترتب شبكة الاستدلال **inference network**، وهي امتداد لأنموذج الاحتمالات، الوثائق طبقيا، وفقا لتراجع الاحتمالات، وبذلك يمكنها أن تفي بحاجة المستفيد من المعلومات، لا باحتمالات أن تكون المعلومات صالحة، كما في أنموذج الاحتمالات الأساس (**Spärck Jones & Willett, 1997**).

ومن الممكن أن نجد امتدادات إضافية للنماذج الأساس لاسترجاع المعلومات، في التقسيم الهرمي لنماذج استرجاع المعلومات، الذي طرحه كل من بايزا - ييتس وريبريو - نتو (**Baeza-Yates and Riberio-Neto (1999**. وسوف نتناول بمزيد من التفصيل، اثنين من الامتدادات الشهيرة، وهما المنطق البوليني الممتد، ومجموعة الغموض أو التشويش، في القسمين التاليين.

أنموذج المنطق البوليني الممتد:

من عيوب أنموذج المنطق البوليني، كما سبق أن ذكرنا، عدم تقبل أوزان المصطلحات، كما أن من مظاهر ضعف أنموذج الفضاء المتجهي عدم القدرة على التعبير عن العلاقات المنطقية البولينية. وقد تطور أنموذج المنطق البوليني الممتد لاستيعاب المصطلحات الموزونة، مع المحافظة على البنية البولينية. وعلى الرغم من أن بوكستاين **Bookstein** وآخرين قد أجروا أيضا بحوثا في هذا الموضوع (**Bookstein, 1978** على سبيل المثال) فإن هاري وو **Harry Wu** هو من أدخل، تحت إشراف جيرارد سالتون **Gerard Salton**، أنموذج المنطق البوليني الممتد، بينما كان يكتب أطروحته (**Salton, Fox & Wu, 1983; Wu, 1981**).

وفي أنموذج المنطق البوليني الممتد، يتم تخصيص أو تحديد أوزان المصطلحات بناء على أي ربط بين المحددات التالية؛ وهي التقارب، والموقع، والتردد، والصلة المتوقعة. وبفضل خاصية وزن المصطلحات، يمكن للأنموذج أن يقدم مخرجات مرتبة

طبقيا، وبذلك يتحكم في عدد الوثائق المسترجعة. وفي الوقت نفسه تتم المحافظة على مزايا تكوين الاستفسار البوليني المنضبط.

إلا أن أنموذج المنطق البوليني الممتد، لم يطبق على نطاق واسع في استرجاع المعلومات، نظرا للأسباب التالية: أولا، من الصعب تحديد أوزان المصطلحات على نحو فعال ودقيق، نظرا للقضايا نفسها التي أوجزناها في القسم ٧/ ٣/ ٢، فيما يتعلق بأوجه قصور أنموذج الفضاء المتجهي. ثانيا، يمكن للاستفسارات المتناظرة منطقيا، أن تفشل في تقديم النتائج نفسها، عندما يتم تحديد أوزان مختلفة لمصطلحات الاستفسار (**Korfhage, 1997**). وعادة ما يسترجع المصطلح ثقيل الوزن عددا كبيرا من الوثائق. إلا أن هذا الأنموذج يحاول التغلب على ما يكتنف أنموذجي المنطق البوليني والفضاء المتجهي من قصور، وذلك بالجمع بين مظاهر قوتهما. وتدعم بعض نظم استرجاع الإنترنت (جوجل على سبيل المثال) البحث بالمنطق البوليني الممتد.

أنموذج الفئات المشوشة:

يهدف أنموذج الفئات المشوشة **fuzzy set**، الذي اقترحه في البداية زاده **Zade (1965)**، إلى التغلب على أحد مظاهر قصور أنموذج المنطق البوليني (وهو عدم القدرة على التعبير عن الصلاحية الجزئية لنتائج البحث)، وذلك بتعميم نظرية الفئات التقليدية. ففي الفئة التقليدية، إما أن يكون المقصود أو الهدف في الفئة أو ليس في الفئة. وكذلك الحال أيضا فإن الوثيقة إما أن تكون صالحة أو مرتبطة، وإما غير صالحة بالنسبة لاستفسار معين، في النظام البوليني. ويعزل هذا الحد الفاصل أعضاء الفئة عن غير الأعضاء، أي الوثائق الصالحة عن الوثائق غير الصالحة. إلا أن مثل هذا الخط الفاصل لا وجود له تقريبا في استرجاع المعلومات، نظرا لأن النظام والمستفيد، لا يستطيعان في غالب الأحيان أن يحددا بدقة، ما إذا كانت الوثيقة صالحة لاستفسار بعينه أم لا (**Korfhage, 1997**)، وإنما يمكن للصلاحية الجزئية أن تكون أكثر دقة في التعبير عن الحكم.

ويُعرف مفهوم العضوية الجزئية- المقابل لفكرة الصلاحية الجزئية، اللازم للارتفاع بمستوى أنموذج المنطق البوليني- بنظرية الفئات المشوشة. ويفترض هذا الأنموذج أن تكون الوثائق والاستفسارات المشوشة لا وجود لها، إلا أنه من الممكن إصدار حكم مشوش. ويتم التحقق من درجات العضوية بالنسبة للفئات المشوشة، فيما يتراوح بين ٠.٠ و ١.٠، حيث يدل ١.٠ (الواحد الصحيح) على العضوية الكاملة. ويمكن للمقصود أن ينتمي إلى فئة ما جزئيا، عندما تتراوح درجة عضويته بين ٠.٠ و ١.٠ وبذلك يكون الحد الذي يفصل الأعضاء عن غير الأعضاء في الفئة مشوشا.

ففئة الطلبة المتربعين على القمة أو المتفوقين، على سبيل المثال، يتم انتقاؤها من الحاصلين على مرتبة الشرف. ويمكن للانتقاء أن يتم بطريقتين؛ أولاهما اتباع نظرية الفئات التقليدية بتعريف "الطلبة المتفوقين" على وجه التحديد، بوصفهم أولئك الحاصلون على المعدل التراكمي ٣.٩ فأكثر. ومن شأن أي طالب تنطبق عليه المواصفات أن يصبح عضوا في فئة الحاصلين على مرتبة الشرف. ولا يسمح لجميع الطلبة الحاصلين على معدل تراكمي أدنى من ٣.٩ أن يدخلوا أو يصبحوا أعضاء في هذه الفئة، حتى وإن كان من الممكن لبعضهم أن يكون حاصلا على معدل تراكمي ٣.٨٩. وفي مقابل ذلك، فإن الطريقة الأخرى لانتقاء الطلبة المتفوقين لمرتبة الشرف، تستخدم نظرية الفئات المشوشة، وذلك بتحديد درجة عضوية لكل طالب. ومن ثم فإن الطلبة المتفوقين يتم تحديدهم بدرجات العضوية التي يحصلون عليها؛ فأولئك الذين حصلوا على معدل تراكمي مقداره ٣.٩ يمكن منحهم درجة العضوية ١.٠، بينما يحصل الطلبة الذين يبلغ معدلهم التراكمي ٣.٥ فأقل على درجة عضوية تقترب من ٠.٠، أما الطلبة الحاصلون على معدل تراكمي في الوسط، وليكن ٣.٨، فيحصلون على درجة عضوية ٠.٨ على سبيل المثال. ومن ثم، فإن الطلبة الذين لم يحصلوا على معدل تراكمي ٣.٩، يمكن أن يظلوا ضمن فئة الطلبة المتفوقين، إذ تدل درجات عضويتهم على ارتباطهم بالفئة.

وفي تطبيق نظرية الفئات المشوشة في استرجاع المعلومات، يمكن الحكم على صلاحية الوثائق، بوصفها صالحة أو غير صالحة، فقط، كما يحدث في نظام المنطق البوليني، وعوضا عن ذلك يمكن تحديد درجة عضوية للوثائق للدلالة على مدى قوة انتمائها إلى فئة الوثائق الصالحة. وتتقرر درجات العضوية بالنسبة للفئة المشوشة من الوثائق، من جانب المكشف في أثناء عملية التكشيف **(Bookstein, 1985)**.

وتشمل مظاهر قوة أنموذج الفئات المشوشة التخفيف من حدة القيود الصارمة في أنموذج المنطق البوليني، أي أن تكون الوثائق إما صالحة للاستفسار وإما غير صالحة، دون السماح بالصلاحية الجزئية. وفي تحديد درجة العضوية يمكن لنظام الفئات المشوشة لاسترجاع المعلومات، أن يوفر مخرجات مرتبة طبقيا، وفقا لتناقص الصلاحية، بحيث يكون بإمكان المستفيد الاتجاه نحو النظر في النتائج المتربعة على القمة. وفضلا عن ذلك، فإنه تتم المحافظة على بنيان الاستفسار البوليني في الأنموذج، للتعبير عن العلاقات المنطقية. وتصلح النظم القائمة على أنموذج الفئات المشوشة للاسترجاع الاستكشافي.

بيد أن نظام استرجاع المعلومات القائم على الفئات المشوشة لا يتسم بالمرونة كما ينبغي، نظرا لأنه لا يحدد أوزانا لمصطلحات الاستفسارات في مقابل مصطلحات الوثائق. ولهذا، فإن قيم استرجاع الوثائق تتوقف فقط على قيم مصطلحات الوثائق **(Salton, 1989)**. وتتعلق هذه العبارة بتداعيات المحافظة على الإطار البوليني، في أنموذج الفئات المشوشة؛ ففي مثل عمليات البحث البولينية هذه، التي تستعمل المعامل "أو" **Or** (مثل أ أو ب أو ج) من المفترض في الوثيقة **D1** المرتبطة بمصطلح استفسار واحد فقط (أي أ) أن تكون مهمة، شأنها في ذلك شأن الوثيقة **D2** التي تشتمل على جميع مصطلحات الاستفسار (أي أ، وب، وج). ومما لا شك فيه أن قيمة أو صلاحية الوثيقة **D1** يتم الحكم عليها بناء على مصطلح واحد فقط (أي أ) في هذه

الحالة، نظرا لأن مصطلحات الاستفسار لم يتم وزنها. أما فيما يتعلق بالاستفسارات التي تستعمل المعامل "و" AND (مثل أ و ب وج) فإن الوثيقة D1 التي تشتمل على جميع مصطلحات الاستفسار فيما عدا واحد (أي أ و ب) تعد بلا طائل، أي غير صالحة، شأنها في ذلك شأن الوثيقة D2 التي لا تشتمل على أي من مصطلحات الاستفسار على الإطلاق.

وفضلا عن ذلك، فإنه بالمقارنة بأنموذج الفضاء المتجهي، فإن أنموذج الفئات المشوشة لا يوفر أي وسيلة للتعبير عن الاستفسارات. وفي مقابل أنموذج الاحتمالات، فإن أنموذج الفئات المشوشة أقل تماسكا من الناحية النظرية على الأقل. ولا يطبق هذا الأنموذج سوى في حالات متفرقة نادرة، في مجال استرجاع المعلومات.

نماذج استرجاع المعلومات: نظرة إضافية:

كما بينا في بداية هذا الفصل، فإن نماذج استرجاع المعلومات والامتدادات التي أوضحناها حتى الآن، ترتبط بالنظم في المقام الأول. وترتبط نماذج أخرى لاسترجاع المعلومات بمحددات مختلفة؛ فالأنموذج المعرفي، على سبيل المثال، يضع في الحسبان عامل المستفيد في استرجاع المعلومات، وسوف نتناول هذا الأنموذج في الفصل العاشر.

مراجعة للنماذج الرئيسة لاسترجاع المعلومات:

قدمنا في الأقسام السابقة معالجات سردية للنماذج الثلاثة الرئيسة لاسترجاع المعلومات، وهي المنطق البوليني، والفضاء المتجهي، والاحتمالات. وقد قصد بالجدول رقم ٧/ ١ تلخيص بعض الخصائص المهمة لهذه النماذج.

الجدول رقم ٧/ ١ خصائص النماذج الثلاثة لاسترجاع المعلومات

الأنموذج / الخاصية	المنطق البوليني	الفضاء المتجهي	الاحتمالات
المنطق البوليني	نعم		
الـوزن		نعم	نعم
الترتيب الطبقي		نعم	نعم
معيار المضاهاة	وجود المصطلح	مسافة المتجه	تردد المصطلحات
الخاصية الفريدة		تلقيم الصلاحية المرتد	

ونعيد النظر في النماذج الثلاثة الواردة في الجدول رقم ٧/ ١، من خمسة جوانب مختلفة: (١) هل يحظى المنطق البوليني بالدعم؟ (٢) هل وزن المصطلحات متوافر؟ (٣) هل ترتب المخرجات طبقيا؟ (٤) أي معايير المضاهاة تطبق للتحقق من التشابه بين الاستفسارات والوثائق؟ (٥) هل هناك خاصية فريدة؟ وتبين هذه الجوانب الخمسة، إلى حد بعيد، مظاهر قوة نماذج استرجاع المعلومات ومواطن ضعفها. فتوافر إمكانية البحث البوليني، على سبيل المثال، تتيح للمستفيد القدرة على تكوين استفسارات محكمة. ومن ناحية أخرى، يمكن للخاصية نفسها (أي المنطق البوليني) أن تجعل نظام استرجاع المعلومات غير متعاطف مع المستفيد، نظرا للصعوبات التي ينطوي عليها إجراء عمليات البحث البولينية.

ويبدو أنموذج المنطق البوليني هو الأضعف من بين النماذج الثلاثة، الواردة في الجدول رقم ٧/ ١؛ فهو لا يدعم سوى البحث البوليني فقط، وتتوقف المضاهاة على

وجود المصطلح في النظام. إلا أن هذا الأنموذج يستخدم بكثافة، ويطبق على نطاق واسع في استرجاع المعلومات. أما الأنموذجان الآخران فيبدوان متشابهين على السطح؛ إذ إنهما، في شكلهما الأصلي لا يشتملان على المنطق البوليني، إلا أنهما يكفلان تقنيات الوزن والترتيب الطبقي. إلا أن المعايير التي تستخدم من جانب أنموذجي الفضاء المتجهي والاحتمالات، لوزن المصطلحات، وترتيب المخرجات طبقيا، وقياس التشابه، مختلفة. يضاف إلى ذلك أن تلقيم الصلاحية المرتد قد أضيف كإمكانية استرجاع فريدة في أنموذج الفضاء المتجهي. وقد بذلت جهود لاستعمال أي من أنموذجي الفضاء المتجهي والاحتمالات، لتطوير نظم أفضل لاسترجاع المعلومات. إلا أن مثل هذه النظم لم تكن قادرة على أن تتفوق في أدائها، على النظم القائمة على أنموذج المنطق البوليني **(Korfhage, 1997)**.

نماذج استرجاع المعلومات في مقابل تقنيات الاسترجاع:

قدمنا وناقشنا في الفصل الخامس مختلف تقنيات الاسترجاع، فما العلاقة إذن بين نماذج استرجاع المعلومات وتقنيات الاسترجاع؟ وبينما لا نجد هناك أي ارتباط دقيق، أو مباشر صريح بين الطرفين، فإن بعض تقنيات الاسترجاع ترتبط بوضوح بنماذج استرجاع المعلومات التي اشتقت منها؛ فالبحث البوليني، على سبيل المثال، تطبيق ولاشك لأنموذج المنطق البوليني. كذلك تحظى توسعة الاستفسارات، وتلقيم الصلاحية المرتد على وجه الخصوص، بالدعم من جانب أنموذج الفضاء المتجهي. ومن الممكن للبحث بالمصطلحات الموزونة أن يستند إلى خوارزميات تطورت في أنموذج الاحتمالات، أو أي نماذج أخرى لاسترجاع المعلومات، كالمنطق البوليني الممتد على سبيل المثال.

وإلى جانب تقنيات الاسترجاع تلك، ذات الأصول المعروفة بوضوح، تستند بقية التقنيات أيضا إلى نماذج استرجاع المعلومات؛ فالبحث بالتقارب على سبيل المثال،

تمتد جذوره في الأنموذج البوليني الممتد. وفضلا عن ذلك، فإن تقنيات الاسترجاع ليس من الضروري أن تشتق من أنموذج استرجاع المعلومات نفسه الذي ينبغي أن يستخدم معها أثناء عملية البحث. ومن الممكن لكل من النماذج والتقنيات أن تستخدم جنبا إلى جنب، بصرف النظر عن مصادرها أو أصولها، إذا ما كان نظام استرجاع المعلومات قد صمم على هذا النحو. وواقع الأمر أن بناء نظم استرجاع المعلومات متعددة النماذج ينبغي أن يحظى بالتشجيع، على النحو المبين في القسم ٧/ ٦/ ٣.

ويمكن للإحاطة بالعلاقة بين نماذج استرجاع المعلومات وتقنيات الاسترجاع أن تساعد في اختيار النظم المناسبة لمهام الاسترجاع. ومن بين الأمثلة الواضحة أن النظام البوليني يمكن أن يقع عليه الاختيار إذا كانت لدى المستفيد الرغبة في إجراء عمليات البحث البوليني. وإذا كان الأمر يتطلب إجراء بحث بناء على وزن المصطلحات، فإنه يتعين على المستفيد اختيار نظام لاسترجاع المعلومات، يستند إلى أنموذج الفضاء المتجهي، على سبيل المثال.

في الطريق إلى نظم استرجاع المعلومات متعددة النماذج:

لنماذج استرجاع المعلومات، كما تناولناها في هذا الفصل، مالها وما عليها. ومن ثم فإن نظم استرجاع المعلومات التي تطبق هذه النماذج، لا يمكنها سوى إنجاز مهام استرجاع معينة فقط. ولاستغلال مزايا ما لمختلف نماذج استرجاع المعلومات من مظاهر قوة، ينبغي بناء نظم استرجاع المعلومات بحيث تتضمن خصائص من نماذج استرجاع معلومات متعددة. وقد عبر فرانتس **Frants** وآخرون عن فكرة مماثلة **(Frants, et al., 1999)**، وإن كانوا يسمون النظم التي تتبع مبادئ متعددة لاسترجاع المعلومات "النظم متعددة الطبعات **multiversion systems**".

وتدل الممارسات الراهنة في مجال استرجاع المعلومات، على أن المنطق البوليني هو الأنموذج الوحيد الذي يطبق في جميع نظم استرجاع المعلومات تقريبا، بينما تطبق النماذج الأخرى تدريجيا في نظم كنظم استرجاع الإنترنت. ويمكن لاستفسارات المعلومات أن تكون بدائية جدا، كما يمكن أن تكون في غاية التطور. وينبغي لنظم استرجاع المعلومات اللازمة لمعالجة هذه الاستفسارات، أن تصمم تبعا لذلك، باتباع طرق متعددة النماذج. ونظم استرجاع المعلومات متعددة النماذج، في تزايد مستمر على الإنترنت، التي في سبيلها لأن تصبح منصة رئيسة لنظم استرجاع المعلومات.

والقضية الأخيرة فيما يتعلق بنظم استرجاع المعلومات متعددة النماذج، هي كيف يمكن لمختلف نماذج استرجاع المعلومات أن تتكامل. ومن الممكن الاسترشاد بمشروعات مثل مؤتمرات استرجاع النصوص (ترِك **Text Retrieval (TREC Conferences** في هذا الصدد. كذلك ينبغي أيضا التعرف على المستفيد المستهدف من نظام استرجاع المعلومات، من أجل التحقق من احتياجاته المعلوماتية على نحو أفضل في هذا العصر الرقمي.

المراجع

Baeza-Yates, Ricardo, and Riberio-Neto, Berthier. (1999). *Modern information retrieval.* New York: ACM Book Press.

Belkin, Nicholas J., and Croft, W. Bruce. (1987). Retrieval techniques. *Annual Review of Information Science and Technology,* 22, 109-145.

Bookstein, Abraham. (1978). On the perils of merging Boolean and weighted retrieval systems. *Journal of the American Society for Information Science,* 29(3), 156-158.

Bookstein, Abraham. (1985). Probability and fuzzy-set applications to information retrieval. *Annual Review of Information Science and Technology,* 20, 117-151.

Chowdhury, Gobinda G. (1999). *Introduction to modern information retrieval.* London: Library Association Publishing.

Cooper, W. S. (1988). Getting beyond Boole. *Information Processing & Management,* 24, 243-248.

Frants, Valery I., et al. (1999). Boolean search: Current state and perspectives. *Journal of the American Soceity for Information Science,* 50(1), 86-95.

Ingwersen, Peter, and Järvelin, Kalervo. (2005). *The turn: Integrating of information seeking and retrieval in context.* Dordrecht, The Netherlands: Springer.

Korfhage, Robert R. (1997). *Information storage and retrieval.* New York: John Wiley & Sons.

Kowalski, Gerald. (1997). *Information retrieval systems: Theory and implementation.* Boston: Kluwer Academic Publishers.

Lesk, Michael. (1997). *Practical digital libraries: Books, bytes and bucks.* San Franciso: Morgan Kaufmann.

Maron, M. E. and Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. *Journal of the ACM,* 7, 216-244.

Robertson, S. E. (1977). The probability ranking principle in IR. *Journal of Documentation,* 33, 294-304.

Robertson, S. E., and Spärck Jones, K. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science,* 27, 129-146.

Salton, Gerard. (1968). *Automatic information organization and retrieval.* New York: McGraw-Hill.

Salton, Gerard. (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by computer.* New York: Addison-Wesley.

Salton, G., Fox, E. A., and Wu, H. (1983). Extended Boolean information retrieval. *Communications of the ACM,* 26, 1022-1036.

Salton, G., Wong, A., and Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM,* 18, 613-620.

Soukhanov, Anne H., et al. (Eds.). (1984). *Webster's II new Riverside university dictionary.* Boston: Riverside Publishing Co.

Spärck Jones, Karen, and Willett, Peter. (Eds.). (1997). *Readings in information retrieval.* San Francisco: Morgan Kaufmann.

Wu, Harry. (1981). *On query formulation in information retrieval.* Unpublished doctoral dissertation, Cornell University, Ithaca, NY.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control,* 8, 338-353.

الفصل الثامن

نظم استرجاع المعلومات

يتم إنجاز مهام استرجاع المعلومات في إطار نظم استرجاع المعلومات. وقد تطورت أنواع مختلفة من نظم استرجاع المعلومات، منذ خمسينيات القرن العشرين، لتلبية فئات مختلفة من الاحتياجات المعلوماتية. ونظم الخط المباشر، ونظم الأسطوانات المكتنزة **CD-ROM**، والفهارس المتاحة للجمهور على الخط المباشر (أوباك **OPACs**)، ونظم استرجاع معلومات الإنترنت، هي الفئات الأربع الرئيسة لنظم استرجاع المعلومات، التي تخدم المستفيدين، بقدرات متفاوتة، لتلبية طلباتهم المعلوماتية. ونتناول في هذا الفصل بالدرس، كل نوع من النظم، لبيان مقوماتها ووظائفها وإمكاناتها في استرجاع المعلومات.

نظم الخط المباشر: النظم الرائدة لاسترجاع المعلومات:

غالبا ما يتم اختصار نظم الخط المباشر **online** لاسترجاع المعلومات، التي تسمى أيضا نظم الخط المباشر المهنية، كنظم الخط المباشر، أو مراصد بيانات الخط المباشر. وتعامل كلمة مرصد البيانات **database** بوصفها مرادفا للنظام في الحالة الأخيرة. وهذه هي أول أنواع نظم استرجاع المعلومات، التي استخدمت تقنيات الحاسبات. وديالوج **DIALOG** والمدلاين **MEDLINE** مثالان لنظم الخط المباشر لاسترجاع المعلومات(*).

(*) الواقع أن ديالوج نظام مضيف **host**، يهيئ فرص التعامل مع مراصد بيانات متعددة على الخط المباشر. أما المدلاين فهو الشكل الذي اتخذه نظام تحليل الإنتاج الفكري الطبي (المدلرز **MEDLARS**)، عندما أتيح التعامل معه على الخط المباشر. (المترجم)

وتكفل نظم الخط المباشر للمستفيد، بمساعدة الحاسب وتقنيات الاتصالات بعيدة المدى، فرصة البحث في مراصد البيانات القائمة في أماكن نائية. وكانت النظم في البداية تدعم التعامل على دفعات **batch processing** فقط (أي أن طلب البحث لم يكن يعالج فورا بمجرد تقديمه، وإنما كان يعالج فيما بعد، عندما يتجمع عدد معين من الطلبات)، ثم أدخلت فيما بعد مقومات التفاعل الفوري بين المستفيد والنظام. وقد مر تطور نظم الخط المباشر في ثلاث مراحل (**Bourne, 1980**):

- دراسات الجدوى والمشروعات الريادية: خمسينيات القرن العشرين.
- الإنتاج بمجتمعات محدودة من المستفيدين: ستينيات القرن العشرين.
- خدمات استرجاع المعلومات الوطنية أو الدولية: من سبعينيات القرن العشرين حتى الوقت الحاضر.

وقد قدم هان (**Hahn (1996**) معالجة مستفيضة للرواد الذين شاركوا في تطوير نظم الخط المباشر. وتتبع بورن وهان (**Bourne and Hahn (2003**) تطورات نظم الخط المباشر بين عامي ١٩٦٣ و١٩٧٦. ففي أعقاب التطورات والتحسينات التي تحققت منذ خمسينيات القرن العشرين، أصبحت نظم الخط المباشر فئة متميزة من نظم استرجاع المعلومات، في العصر الرقمي.

خصائص نظم الخط المباشر لاسترجاع المعلومات:

لقد كانت نظم الخط المباشر تجمع في المقام الأول، المعلومات الوراقية وبعض المعلومات العددية قبل ثمانينيات القرن العشرين، ثم أصبحت بعد ذلك تشمل معلومات النصوص الكاملة. وقلما ينظر إلى معلومات الوسائط المتعددة في هذا النوع من نظم الاسترجاع. ويتم اختيار المعلومات التي تتضمنها نظم الخط المباشر، وتنظيمها بواسطة اختصاصين (كالمكشفين مثلا). وتستعمل اللغات المقيدة بكثافة،

لأغراض تنظيم المعلومات واسترجاعها، كما يحظى البحث بالكلمات المفتاحية بالدعم في الوقت نفسه. وكانت لغة الأوامر **command** هي القاعدة بالنسبة للمستفيدين، للتفاعل مع نظم الخط المباشر، إلى أن تم إدخال الاختيار من قوائم الخيارات، في ثمانينيات القرن العشرين. ثم بدأ استخدام واجهات التعامل التصويرية **graphical** تدريجيا، في تسعينيات القرن العشرين، وخصوصا عندما بدأت نظم الخط المباشر تستخدم العنكبوتية العالمية بوصفها منصة لها.

وقد اتبعت فئات متعددة من تقنيات الاسترجاع، ونماذج استرجاع المعلومات، في نظم الخط المباشر. والتقنيات الأساس للاسترجاع، كالبحث البوليني، والبحث بمراعاة شكل الحرف الأول، والبتر،و البحث بالتقارب، والبحث بالحقول، متاحة في جميع نظم الخط المباشر تقريبا. أما تقنيات الاسترجاع المتطورة، كالبحث الموزون، والبحث المشوش **fuzzy**، وتوسعة الاستفسارات، وإن كانت لا تحظى بالدعم في جميع نظم الخط المباشر، فإنها يمكن أن تتوافر في بعض التطبيقات. وواقع الأمر أن نظم الخط المباشر كانت، فيما قبل منتصف تسعينيات القرن العشرين، بمثابة بيئات اختبار بالنسبة للبحث والتطوير في استرجاع المعلومات. كما كانت هذه النظم أيضا تقوم مقام نوافذ العرض بالنسبة لتقنيات استرجاع المعلومات، قبل ظهور نظم استرجاع الإنترنت؛ فقد كان ديالوج على سبيل المثال، يكفل المخرجات المرتبة طبقيا، باستعمال الأمر **RANK** في عام ١٩٩٣ **(Basch, 1993)**.

وفضلا عن ذلك، يمكن النظر إلى نظم الخط المباشر بوصها مختبرات لاكتساب وتنمية مهارات استرجاع المعلومات، لعدة أسباب؛ أولها أن النظام يتم بناؤه بحيث يكفل الممارسة والمعالجة المنضبطة. وهناك بعض نظم الخط المباشر التي توفر مقومات خاصة (كملفات وورش عمل أونتاب ديالوج **DIALOG's ONTAP**(*))

(*) الاسم المختصر لبرمجيات التدريب على الخط المباشر **Online Training and Practice**. (المترجم)

للمساعدة في تدريب المستفيدين. ثانيا، كانت نظم الخط المباشر تمثل النظم الإلكترونية الوحيدة لاسترجاع المعلومات المتاحة في ذلك الوقت. ثالثا، من الممكن اختبار وتجريب فئات متنوعة من مهارات الاسترجاع في نظم الخط المباشر. إلا أن السبب الرابع يأتي في الجانب السلبي؛ فمن المفترض أن يحصل المستفيدون من نظم الخط المباشر حتما على التدريب، قبل الإفادة من هذه النظم أو التعامل معها. فالنظام ليس حدسيا أو بدهيا، ولا يمكن الإلمام به ببساطة بالتجربة والخطأ. ونتيجة لذلك، فإن المستفيدين النهائيين، في البحث على الخط المباشر اختصاصيو معلومات بوجه عام، يقومون بدور وسطاء البحث، بين النظام والمستفيد الذي يحتاج إلى المعلومات.

نظم الخط المباشر واسترجاع المعلومات:

من بين الأنواع الأربعة لنظم استرجاع المعلومات التي تحققنا منها في هذا الفصل، تتمتع نظم الخط المباشر بأطول تاريخ، كما أن لها أقوى تأثير في استرجاع المعلومات. فهذه النظم ترسي ضمنيا المواصفات المعيارية التي تلتزم بها النظم الأخرى. وحيثما يظهر نوع جديد من نظم استرجاع المعلومات، دائما ما يقارنه المستفيدون المحتملون بنظم الخط المباشر، ليروا إلى أي حد يمكن للوافد الجديد أن يعمل بكفاءة في المجال؛ فقد تناول نال-جاكوبوفيتز وتينوبير **Nahl-Jakobovits and Tenoir (1992)**، على سبيل المثال بالدرس، جوانب زمن الاستجابة، والتغطية، والمحتوى، والتكلفة، في كل من طبعتي المستخلصات السيكولوجية **Psychological Abstracts** والمستخلصات الاجتماعية **Sociological Abstracts**، على الأسطوانات المكتنزة وعلى الخط المباشر. وقارن هيلدرث **Hildreth (1988)** الفهارس المتاحة للجمهور على الخط المباشر بنظم استرجاع المعلومات على الخط المباشر، واستكشف احتمالات إدخال خصائص نظم الخط المباشر، في الفهارس المتاحة للجمهور على الخط المباشر. كذلك قارنت تشو **Chu**

(1998) نظم استرجاع الإنترنت بنظم الخط المباشر، في جوانب كبناء مراصد البيانات، وإمكانيات البحث، والأداء في الاسترجاع، وخيارات المخرجات، وجهد المستفيد.

ويبدو أن نظم الخط المباشر قد أثبتت مكانتها، بوصفها النظم التي يقاس عليها، في استرجاع المعلومات. إلا أنه كما سبق أن ذكرنا، فإن نظم الخط المباشر، أبعد ما تكون عن الكمال؛ فالتكلفة المرتفعة، وغياب التعاطف مع المستفيد، من الخصائص التي غالبا ما ترتبط بهذه الفئة من النظم. ومن ناحية أخرى، فإن الدور الرائد الذي نهضت به نظم الخط المباشر، في استرجاع المعلومات، يحظى أيضا بالتقدير على نطاق واسع.

نظم الأسطوانات المكتنزة: وسط مختلف لنظم استرجاع المعلومات:

نشأت نظم الأسطوانات المكتنزة عن تطبيق تقنيات هذه الأسطوانات في استرجاع المعلومات، وحققت إلى أبعد حد، جزءا مما صممت لتحقيقه في مجال استرجاع المعلومات، مسايرة لما حققت تقنيات العنكبوتية العالمية من نضج. ونتناول هذه النظم في هذا الفصل، من أجل استكمال صورة مشهد نظم استرجاع المعلومات، في المقام الأول. وعادة ما يتم البحث في نظم الأسطوانات المكتنزة محليا، إذ لا تعتمد هذه النظم على الاتصالات بعيدة المدى، للتعامل معها ما لم تكن النظم متشابكة. وكان مرصد بيانات الإنتاج الفكري السيكولوجي **PsycLIT**، الذي تتبناه مؤسسة سيلفر بلاتر **SliverPlatter** مثالا لنظم الأسطوانات المكتنزة.

ومن الممكن النظر إلى نظم الأسطوانات المكتنزة، على نحو ما، بوصفها نظم خط مباشر، على وسط الأسطوانات المكتنزة، نظرا لأن هذين النوعين من نظم

استرجاع المعلومات يتقاسمان كثيرا من الخصائص، ولأن نظمَ الأسطوانات المكتنزة مصممة على غرار نظم الخط المباشر، على أكثر من نحو. وفضلا عن ذلك فإن نظم الأسطوانات المكتنزة لم تطبق على نطاق واسع، إلا في ثمانينيات القرن العشرين، عندما أصبحت نظم الخط المباشر بالغة التأثير في المجال فعلا.

ولا تبدو هناك مراحل واضحة في تطور نظم الأسطوانات المكتنزة. ومن بين الأسباب الرئيسة لهذه الظاهرة، أن هذه النظم لم تبدأ من الصفر، وإنما نشأت بوصفها مزاوجة بين الأسطوانات المكتنزة الناشئة، ونظم الخط المباشر الناضجة، باستخدام الأولى كوسط اختزان جديد. ولهذا فإن الخصائص الفريدة لنظم الأسطوانات المكتنزة تتأثر إلى حد بعيد، بخصائص الأسطوانات المكتنزة كوسط.

خصائص نظم الأسطوانات المكتنزة:

لا تزال المعلومات الوراقية، والعددية أو الإحصائية، والنصوص الكاملة، هي مصادر المعلومات الغالبة، التي يتم تجهيزها في نظم الأسطوانات المكتنزة. إلا أن زيادة في طاقة اختزان معلومات الوسائط المتعددة، حدثت عندما أصبحت تقنيات الأسطوانات المكتنزة أكثر قدرة من نظم الخط المباشر، في معالجة مثل هذه الأنواع من المعلومات. ولا تزال نظم الأسطوانات المكتنزة متعددة الوسائط، هي المستخدمة في بعض الظروف (كالموسوعات على سبيل المثال). وتستخدم اللغات المقيدة لأغراض تنظيم المعلومات واسترجاعها، مرتبطة باللغة الطبيعية. ولا يزال التدخل البشري كثيفا، في استرجاع معلومات الأسطوانات المكتنزة. وقلما نجد لغة الأوامر في نظم الأسطوانات المكتنزة، إذ يطبق الاختيار من قوائم الخيارات في غالب الأحيان. وقد ظهرت واجهات التعامل التصويرية أول ما ظهرت في بيئة الأسطوانات المكتنزة، كما أن بعض النظم تتبع البنيان الفائق **hyperstructure** في تطبيقاتها.

وتحظى تقنيات الاسترجاع الرئيسة (كالبحث البوليني، والبحث بالتقارب على سبيل المثال) بالدعم في نظم الأسطوانات المكتنزة، أما إمكانيات البحث المتقدم (كالبحث بالوزن والترتيب الطبقي للمخرجات على سبيل المثال) فتطبق على نطاق محدود. وليس من غير المألوف أن نجد أن نظاما واحدا بعينه، قد تم اختزانه على عدة اسطوانات مكتنزة، الأمر الذي يمكن أن يحول دون تطبيق بعض تقنيات الاسترجاع المتقدمة في النظام.

وبفضل سعتها الاختزانية العالية، وحجمها المادي الصغير نسبيا، فإن نظم الأسطوانات المكتنزة قابلة نظريا، للتعامل معها في أكثر من منصة واحدة **portable**، نظرا لإمكان استيعاب عدة أسطوانات لجميع المعلومات اللازمة في مرصد بيانات معين. إلا أن مرصد البيانات وحده، لا يمكن أن يعمل كنظام لاسترجاع المعلومات. ويتطلب تشغيل النظام تجهيزات خاصة. ومن ناحية أخرى، فإنه يمكن أن يكون من الأكثر ملاءمة بالنسبة للمستفيد، أن يبحث في نظم الأسطوانات المكتنزة، نظرا لأنها لا تعتمد على تقنيات الاتصالات بعيدة المدى في تشغيلها. وإجراء عمليات البحث في الأسطوانات المكتنزة أقل تكلفة، نظرا لأنها لا تتحمل رسوم الارتباط أو أي رسوم أخرى بمجرد شراء النظم وتركيبها، ومن ثم فإنه يمكن للمستفيدين إجراء عمليات البحث، دون الاكتراث بما إذا كانت دقات الساعة تصدر أصواتها، ويمكنهم التركيز أكثر على البحث نفسه.

والمستفيد النهائي لا الوسيط، هو من يقوم بإجراء معظم عمليات ا لبحث في بيئة الأسطوانات المكتنزة. ومن بين أسباب ذلك أنه وفقا لسياسة الرسوم الثابتة، والتعامل غير المحدود، فإن نظم الأسطوانات المكتنزة لا تفرض عادة رسوما إضافية على المستفيد النهائي. أما السبب الثاني فهو أن واجهة تعامل نظم الأسطوانات المكتنزة، أكثر تعاطفا ومراعاة لظروف المستفيد، من نظم الخط المباشر، نتيجة لاستخدام قوائم الخيارات

والواجهات التصويرية. ثالثا، تحث نظم الأسطوانات المكتنزة، وتشجع على التصفح، نظرا لأن تكلفة الخط المباشر لا تشكل عبئا بالنسبة للمستفيد النهائي في هذه الحالة. وكما تبين لنا في الفصل السادس، فإن التصفح والبحث هما النهجان الرئيسان للاسترجاع، والأول هو المفضل في بعض الظروف. وبالمقارنة فإن التصفح في جلسات الاسترجاع على الخط المباشر يبدو مرتفع التكلفة.

إلا أن نظم الأسطوانات المكتنزة يكتنفها بعض القصور في التحديث. إذ ينطوي كل تحديث على إحلال الأسطوانات القائمة، الأمر الذي يحد على نحو ملحوظ من معدل تردد التحديث أو تكراره. وعادة ما يكون معدل تحديث نظم الأسطوانات المكتنزة فصليا أو نصف سنوي، بينما يمكن لتحديث ما يقابل هذه النظم من نظم الخط المباشر أن يتم أسبوعيا أو يوميا، وربما على نحو مستمر (كأخبار بردج للأسواق العالمية **Bridge World Markets**، على ديالوج **DIALOG** على سبيل المثال). ويمكن لسرعة البحث في نظم الأسطوانات المكتنزة أن تكون مشكلة، وخصوصا عندما يكون النظام الواحد موزعا على عدة أسطوانات. كذلك لا تكفل نظم الأسطوانات المكتنزة التعامل عن بعد، ما لم تكن مرتبطة بشبكة. وتسهم كل أوجه القصور هذه في الاستعاضة عن تقنيات الأسطوانات المكتنزة بالعنكبوتية العالمية، في هذا النوع من نظم استرجاع المعلومات.

نظم الأسطوانات المكتنزة واسترجاع المعلومات:

لقد سدت نظم الأسطوانات المكتنزة بعض الثغرات، التي لا تغطيها نظم الخط المباشر، في استرجاع المعلومات، وذلك بالوصول إلى المستفيد النهائي، وتوفير خدمات الاسترجاع لأولئك الذين كان من الممكن لنظم الخط المباشر أن تكون بالنسبة لهم باهظة التكلفة، ومن الصعب التعامل معها. ولنسبة مئوية عالية من نظم

الأسطوانات المكتنزة طبعات مقابلة على الخط المباشر، على الرغم من أن تغطيتها وخصائصها قد لا تكون واحدة.

ويتم تدريجيا الاستعاضة عن الأسطوانات المكتنزة كوسائط للاختزان، بالأسطوانات الرقمية متعددة الاستخدامات **Digital versatile disks (DVD)**، وهي تقنية أسطوانات مكتنزة أكثر حداثة. واعتمادا على طبقتين على كل من سطحيها، تتسع الأسطوانة الرقمية متعددة الاستخدامات لما يبلغ ١٧ جيجا بايت من المعلومات البصرية أو السمعية، أو أي معلومات أخرى. وبالمقارنة فإن الأسطوانة المكتنزة الراهنة بالحجم المادي نفسه، تتسع لأقل من ٧٠٠ ميجا بايت من المعلومات **(Tech Target, 2005)**. ولما كانت الأسطوانات الرقمية متعددة الاستخدامات تمثل تقنية أكثر تطورا، فإنها تحل محل الأسطوانات المكتنزة كوسائط لنظم استرجاع المعلومات القابلة للتعامل مع عدة منصات، من أجل المعلومات متعددة الوسائط. أما الوظائف الأخرى لنظم الأسطوانات المكتنزة، فتتكفل بها تدريجيا تقنيات العنكبوتية العالمية.

الفهارس المتاحة للجمهور على الخط المباشر: الفهارس الإلكترونية للمكتبات كنظم لاسترجاع المعلومات:

الفهارس المتاحة للجمهور على الخط المباشر **OPACs**، فهارس تقليدية محملة على وسط مختلف **(Malinconico, 1984)**. وفيما قبل ثمانينيات القرن العشرين، لم يكن هناك سوى بعض الجهود الرائدة، كنظام التحكم في المكتبة **Library Control System** بجامعة ولاية أوهايو **State University of Ohio**. إلا أنه منذ عام ١٩٨٠ تم تركيب نماذج مبدئية، ونظم عاملة للفهارس المتاحة للجمهور على الخط المباشر، في عدد من المكتبات، ينمو باطراد **(Hildreth, 1985)**. وهناك في أيامنا هذه قليل من المكتبات التي لا يوجد بها فهرس متاح للجمهور على الخط المباشر ضمن نظمها.

والفهارس المتاحة للجمهور على الخط المباشر، سواء كانت قائمة بذاتها أو جزءا من نظام متكامل لميكنة المكتبات مثل هورايزون **Horizon**، من سرسي داينكس **SirsiDynix**، من نتائج ميكنة المكتبات. وقد تم تطوير الفهارس المتاحة للجمهور على الخط المباشر، في البداية بواسطة كل من المتعهدين وفرق العمل بالمكتبات. وقد توقف الخيار الأخير، ولم يعد له وجود، عندما أصبح متعهدو نظم ميكنة المكتبات أكثر خبرة واستقراراً. وقد تغيرت واجهات التعامل مع الفهارس المتاحة للجمهور على الخط المباشر من لغة الأوامر إلى قوائم الخيارات، وتعبئة النماذج والتمثيلات التصويرية.

وقد قسم هايلدرث **Hildreth (1984)** الفهارس المتاحة للجمهور على الخط المباشر إلى ثلاثة أجيال، ليرسم خريطة تاريخها الحديث، كما يتكهن بتصميمها المحتمل في المستقبل. ويحاكي الجيل الأول من هذه الفهارس، بشكل أو بآخر، نهج الفهارس البطاقية بالنسبة لمحتوى الملف، والتنظيم، والتعامل بالنسبة للبحث عن الأوعية المعروفة في المقام الأول. ويضيف الجيل الثاني للفهارس المتاحة للجمهور على الخط المباشر تعزيزات في نقاط الوصول الموضوعية، وإمكانيات البحث، وجوانب أخرى. وتشبه فهارس هذا الجيل نظم الخط المباشر من نواح عدة. أما الجيل الثالث من الفهارس المتاحة للجمهور على الخط المباشر، وهو في طريقه الآن لأن يصبح حقيقة واقعة، فيجعل الفارق بين هذا النوع من الفهارس، ونظم الخط المباشر أقل وضوحا، وذلك بتوسعة مدى التعامل؛ فمن الممكن للمستفيد البحث في مراصد البيانات على الخط المباشر، والتعامل مع المصادر الأخرى عن طريق الفهرس المتاح للجمهور على الخط المباشر، الخاص بالمكتبة المحلية بالنسبة للمستفيد. كذلك تحققت أيضا بعض التعزيزات الأخرى (كالتكامل بين نهجي بحث النصوص واللغات الطبيعية واللغات المقيدة على سبيل المثال) تلك التعزيزات التي اقترحها هايلدرث **Hildreth** عام ١٩٨٤ لهذا الجيل من الفهارس، وأصبحت الآن حقيقة واقعة.

ويبدو أن ترقيم أجيال الفهارس المتاحة للجمهور على الخط المباشر، قد توقف عند الجيل الثالث، وعوضا عن ذلك تكتسب فهارس أخرى تتاح على الخط المباشر، تختلف عن تلك الفهارس الخاصة بالأجيال السابقة، أسماء مختلفة كفهارس الجيل التالي وفهارس القرن الحادي والعشرين، على سبيل المثال **(Antelman, Lynema, & Pace, 2006; Hildreth, 1995; Markey, 2007)**. وقد نشأت فهارس الجيل التالي نتيجة لتطويع تنظيم المعلومات واسترجاعها للبيئة الرقمية، وإدخال تقنيات استرجاع أثبتت كفاءتها **(Antelman, Lynema & Pace, 2006** على سبيل المثال). وعادة ما تصبح العنكبوتية العالمية المنصة، بالنسبة لهذا الجيل من الفهارس المتاحة للجمهور على الخط المباشر، الأمر الذي يكفل تحقيق الخصائص المتنوعة التي توفرها العنكبوتية؛ فالفهارس من هذا النوع تكفل على سبيل المثال، للمستفيدين التفاعل مع النظام عن طريق متصفح العنكبوتية، وواجهة التعامل معها. ومن ثم فإن الحاجة لا تدعو لمزيد من التعلم لبيئة الاسترجاع هذه، بالنسبة لمن يفيدون من الفهارس المتاحة للجمهور على الخط المباشر، نظراً لتآلفهم فعلا معها. كذلك يمكن إنشاء روابط بالمصادر الأخرى (كالكتب الإلكترونية ومقالات الدوريات مثلا)، من تسجيلات الفهرس المتاح للجمهور على الخط المباشر، وذلك لأغراض التعامل السلس، بالإضافة إلى الحصول على رؤوس موضوعات بروابط فائقة، في نطاق الفهرس المتاح على الخط المباشر، من أجل تعديل عمليات البحث.

ومن ناحية أخرى دخلت تقنيات الاسترجاع، التي تطورت خصيصا في نظم استرجاع الإنترنت، إلى الفهارس المتاحة على الخط المباشر، تلك الأساليب التي تتضمن التلقيم المرتد الخاص بالاتصال بالموضوع أو الصلاحية، وترتيب نتائج البحث طبقيا، والارتقاء بمستوى البيئة الخاصة بالتصفح، ومراجعة هجاء استفسارات البحث. وغالبا ما يتم الاستشهاد بفهرس الخط المباشر الخاص بمكتبات

جامعة ولاية شمال كارولينا North Carolina (www.lib.ncsu.edu/cotalog) State University Libraries الذي تم تشغيله عام ٢٠٠٦، بوصفه مثالا للجيل التالي من الفهارس المتاحة للجمهور على الخط المباشر (Antelman, Lynema, & Pace, 2006; Markey, 2007). وينطوي هذا الجيل الجديد من الفهارس المتاحة على الخط المباشر، ولاشك، على تحسينات وتعزيزات لم تتوافر في أسلافه، إلا أن كثيرا مما تصور ملامحه هايلدرث Hildreth (1995) وماركي Markey (2007)، في قوائم توقعاتها، بالنسبة لمستقبل نظم الفهارس المتاحة على الخط المباشر، لم يتحقق بعد.

خصائص الفهارس المتاحة للجمهور على الخط المباشر:

تتحقق التغيرات والتطورات كلما انتقلت الفهارس المتاحة على الخط المباشر من جيل إلى آخر. إلا أن هناك خصائص معينة، تبدو فريدة بالنسبة لهذا النوع من الفهارس، بوصفها نوعا من نظم استرجاع المعلومات؛ فهذا الضرب من الفهارس يشتمل أولا على معلومات وراقية حول موارد المكتبات الخاصة بالمؤسسات مختلفة المستويات (المحلية، والإقليمية، والوطنية على سبيل المثال). وعلى الرغم من إنشاء فهرس واحد يتاح على الخط المباشر لمكتبة واحدة، في معظم الأحيان، هناك حالات أخرى يشتمل فيها فهرس واحد على بيانات موارد عدة مكتبات. وفي مقابل ذلك لا تقتصر الأنواع الثلاثة الأخرى من نظم استرجاع المعلومات، في تغطيتها، على الوثائق المتوافرة في مؤسسة واحدة لا أكثر.

ثانيا، يمكن النظر إلى الفهارس المتاحة على الخط المباشر بوصفها امتدادا لتسجيلات الفهرسة التي تقرأ آليا Machine Readable Cataloging (مارك MARC، أو أي مواصفة معيارية للفهرسة) التي عادة ما يتم إعدادها بواسطة اختصاصيي المكتبات، باتباع مجموعة من القواعد والمواصفات المعيارية (كخطط التصنيف وقواعد الفهرسة على سبيل المثال) لمجتمع المستفيدين الذي تخدمه المكتبة.

ولا يمكن لمعالجة المعلومات في الأنواع الأخرى من نظم استرجاع المعلومـات، اتبـاع قواعد توجيهية متطابقة، واضعين في الحسبان مستفيدين مستهدفين بعينهم؛ فـديالوج **DIALOG** وحده، على سبيل المثال، يتسع نظامه المضيف لمئات من مراصد البيانات. ويمكن لكل مرصد للبيانات أن يتم بناؤه وصيانته باستخدام لغة مقيدة مختلفة(*). وفي الوقت الذي يمكن فيه تحديد معالم مجتمع المستفيدين من مكتبة ما، يبدو من الـصعب التحقق بوضوح ممن يمكن أن يفيد من نظام الخط المباشر، وناهيك عن نظم استرجاع الإنترنت، التي تبدو غير قابلة للحصر، دائمة التغير.

ثالثا، تدعم الفهارس المتاحة للجمهور على الخـط المبـاشر، الآن عـلى الأقـل، البحث بالحقل (كالمؤلف أو العنوان مـثلا) والبحـث بـالكلمات المفتاحيـة، والبحـث البوليني، على الرغم من أن المتعاملين مع المكتبات قلما يستخدمون المقومات البولينية، للأسباب التي نوقشت في الفصول السابقة. ويشكل البحث عن وثيقة معروفة بعينها قطاعا لا يستهان به، من إجمالى عمليات البحث في الفهرس المتاح على الخـط المبـاشر، ويرجع ذلك جزئيا إلى أن الفهارس المتاحة على الخط المباشر، من المحتمـل أن تغطـي أوعية "معروفة" بالنسبة للمستفيد، أكثر من نظم الخط المباشر. أما خصائص البحث المتقدم (كالبحث الموزون على سبيل المثال) فإنهـا لا تتـوافر عـادة بالنـسبة للفهـارس المتاحة للجمهور على الخط المباشر، نظرا لأن أداء الاسترجاع المناسب، من الممكن أن يتحقـق حتـى باسـتخدام إمكانيـات البحـث الأسـاس، ويرجـع الفـضل في ذلـك لمجموعات المكتبات المحددة تحديـدا مناسـبا. إلا أنـه فـيما يـسمى الجيـل التـالي مـن الفهارس المتاحة للجمهور على الخط المباشر، غالبا مـا يحظـى التلقـيم المرتـد الخـاص بالصلاحية أو بالاتصال بالموضوع، بالدعم عن طريق مثل هذه الـروابط الفائقـة مـن

(*) للتغلب على هذه المشكلة، وتيسير مهمة المستفيد، طور ديالوج لغة موحدة للتعامل مع مختلـف مراصد البيانات، وهي لغة التحكم في المهام **Job Control Language**. (المترجم)

قبيل "النتائج المماثلة". كذلك يمكن للمستفيدين أيضا البحث في حدود نتائجهم المسترجعة، للعثور على المزيد من الأوعية المحددة، الأمر الذي يعد إحدى الخواص التي أضيفت حديثا لهذا الجيل من نظم الفهارس المتاحة للجمهور على الخط المباشر.

رابعا، يكفل كثير من فهارس الجيل الراهن للفهارس المتاحة على الخط المباشر، الوصول إلى مصادر أخرى، بما في ذلك نظم الخط المباشر. وفي مقابل ذلك لا يمكن لنظم الخط المباشر كفالة مثل هذا الضرب من خدمات البوابات **Gateway**، نظراً لطبيعتها الربحية. واعتمادا على زد ٣٩.٥٠، المواصفة المعيارية الخاصة باسترجاع المعلومات من مصادر مختلفة، باستعمال واجهة تعامل موحدة، يمكن للمستفيدين البحث في فهارس أخرى متاحة للجمهور على الخط المباشر، في أماكن مختلفة. وكما أوضحنا في القسم ٥/ ١/ ٢/ ٤ فإن خدمة العنكبوتية للبحث والاسترجاع **Search / Retrieve Web Service (SRW)**، والبحث والاسترجاع عن طريق العنوان الموحد للمصدر **Search/Retrieve Via URL (SRU)**، يمكنهما معا أن يحلا محل زد ٣٩.٥٠، بكفالة مهام وظيفية مماثلة، ولكن بدرجة تعقد أقل **(Le Van, 2003)**.

خامسا، يشجع تصميم الفهارس المتاحة للجمهور على الخط المباشر على التصفح، إذ يمكن لمستخدمي هذه الفهارس ممارسة التصفح، عادة بنقاط الوصول كالمؤلف والعنوان مثلا، بينما تقدم الأنواع الأخرى من نظم استرجاع المعلومات، النتائج مرتبة زمنيا أو وفقا للترتيب الطبقي للصلاحية أو الاتصال بالموضوع. والتصفح برقم الطلب في الفهارس المتاحة للجمهور على الخط المباشر، مفيد على وجه الخصوص بالنسبة للمستفيد، نظرا لأن ترتيب عرض النتائج، يشبه إلى حد بعيد، الترتيب المادي لمجموعات المكتبات على الأرفف. ومن ثم، فإن المستفيدين من الفهارس المتاحة على الخط المباشر تتاح لهم فرصة العثور على الأوعية المتصلة موضوعيا بجوار بعضها البعض على الأرفف، قبل التوجه إلى الأرفف بأنفسهم.

غير أن الفهارس المتاحة للجمهور على الخط المباشر لا يزال من الصعب التعامل معها، على الرغم من أن معظم هذه الفهارس ينتمي إلى الجيل الثالث. ويرجع بورجمان **Borgman (1986, 1996)** النتيجة النهائية هذه إلى تصميم هذه الفهارس الذي لم ينجح في استيعاب الدروس المستفادة من دراسات استرجاع المعلومات، وإلى الفهم القاصر لسلوك البحث. ولا يزال افتراض قدرة المستفيدين، بلا تدريب، على التعامل مع الفهارس المتاحة للجمهور على الخط المباشر **(Hildreth, 1988)** مفتقرا حتى الآن إلى السند المناسب.

لقد أصبحت الفهارس المتاحة للجمهور على الخط المباشر درة التاج في نظم استرجاع المعلومات، فيما بين بدايات ثمانينيات القرن العشرين وتسعينيات القرن نفسه، عندما كانت تقارن بالفهارس البطاقية، ونظم الخط المباشر، التي يتم التعامل معها مقابل رسوم. وتبدو صلاحية هذه الفهارس في تراجع مطرد، نتيجة لنشأة نظم استرجاع الإنترنت، وارتفاع مستوى كفاءتها، من حيث القدرات الاسترجاعية ومراعاة ظروف المستفيد. ويعبر الإنتاج الفكري الذي صدر بعد العصر الذهبي للفهارس المتاحة للجمهور على الخط المباشر، عن مخاوف حقيقية بشأن وضعها الراهن وظروفها المستقبلية، بينما جوجل **Google**، نظام استرجاع الإنترنت الذي تتزايد شعبيته، لا يحكم سيطرته على البحث في العنكبوتية العالمية فحسب، وإنما يتوسع في خدماته أيضا إلى ما وراء ذلك (مثل جوجل للبحث عن الكتب **Google Book Search** وجوجل العالم **Google Scholar**). ويبدو نجاح جوجل في البحث في العنكبوتية العالمية، بالإضافة إلى توسعه في نطاق نفوذ الفهارس المتاحة على الخط المباشر، المصدر الرئيس لمثل هذه المخاوف **(Marcum, 2006; Markey, 2007)**.

الفهارس المتاحة للجمهور على الخط المباشر واسترجاع المعلومات:

لقد أثير التساؤل حول ما إذا كانت الفهارس المتاحة للجمهور على الخط المباشر، فهارس مكتبات أم نظم استرجاع معلومات على الخط المباشر، في الماضي **(Hildreth, 1985)**. ويمكن للإجابة عن هذا التساؤل الآن، أن تكون الفهارس المتاحة على الخط المباشر نظم استرجاع معلومات بسماتها الخاصة؛ لقد كانت الفهارس المتاحة على الخط المباشر تساعد المستفيدين في العثور على الأوعية في مجموعات المكتبات، وتفيد من الوظائف التي لم تتوافر في أسلافها من الفهارس البطاقية. وفضلاً عن ذلك تقوم الفهارس المتاحة للجمهور على الخط المباشر مقام البوابات التي تفضي إلى النظم الأخرى لاسترجاع المعلومات، بتوفير مقومات الارتباط بها. ولما كانت المنفردات **monographs** هي الفئة الرئيسة لمجموعات المكتبات، فإن الفهارس المتاحة على الخط المباشر، هي أداة الاسترجاع الرئيسة، إن لم تكن الوحيدة للوصول إلى هذه المنفردات. وبدون الفهارس المتاحة على الخط المباشر، يمكن لمجموعات المكتبات أن يكون من الصعب الوصول إليها في هذا العصر الرقمي، نظرا لأن الأنواع الأخرى من نظم استرجاع المعلومات لم تصمم خصيصا لذلك الهدف.

لقد غيرت نشأة خدمة جوجل للبحث عن الكتب **Google Book Search**، كما سبق أن ذكرنا، بل إنها تحدت فعلا الدور المسيطر الذي تنهض به الفهارس المتاحة على الخط المباشر، في العثور على معلومات الكتب. وقد بدأت خدمة جوجل للبحث عن الكتب، التي كانت تسمى في البداية مطبوعات جوجل **Google Print**، عام ٢٠٠٤، لصالح مجموعة من الناشرين (مثل **Blackwell, Cambridge University Press, the University of Chicago Press**)، وبعد شهرين امتدت لمكتبات في عدة مؤسسات (مثل **Harvard, University of Michigan, New York Public Library, Oxford, and Stanford**)، وذلك كمشروع

جوجل للمكتبات (Google, 2008a). ومنذ ذلك الحين وأعداد المشاركين في هذا المشروع في تزايد مطرد. والكتب التي يشملها هذا المشروع قابلة للبحث بعد رقمنتها. وفضلا عن ذلك تتوافر في خدمة جوجل للبحث عن الكتب، خواص كالربط بمراجعات الكتب **Book reviews** والمصادر التي تستشهد بالكتب (كصفحات العنكبوتية العالمية) الخاصة بالكتاب الذي يتم البحث عنه. ومما لاشك فيه، أن خدمة جوجل للبحث عن الكتب، تيسر استرجاع معلومات الكتب، بنهج يختلف عن طريقة الفهارس المتاحة على الخط المباشر. كذلك تدعو خدمة جوجل للبحث عن الكتب، إلى جانب نظامه الخاص بالبحث في العنكبوتية، الفهارس المتاحة على الخط المباشر، لأن تعيد النظر في موقفها ووظائفها، وأدائها في هذا العصر الرقمي لاسترجاع المعلومات (Marcum, 2006).

نظم استرجاع الإنترنت: أحدث أعضاء أسرة نظم استرجاع المعلومات:

يكفل النمو الأسي للإنترنت، للمتعاملين معها القدرة على الوصول إلى كميات هائلة من المعلومات الرقمية، بصرف النظر عن الزمان والمكان. كذلك تتضح في الوقت نفسه، الحاجة إلى المزيد من نظم استرجاع المعلومات، من أجل استرجاع أي "إبر" بعينها من "كومة تبن" الإنترنت. ومن ثم طور الأفراد والشركات، والمؤسسات عددا كبيرا من نظم استرجاع الإنترنت، أو نظم استرجاع المعلومات الخاصة بالإنترنت، للتغلب على مشكلة تفجر المعلومات على الإنترنت.

وكما هو معروف، فإن عبارة محرك البحث في العنكبوتية **web search engine**، أو محرك البحث **search engine** ببساطة، يبدو الاسم الأكثر شيوعا من نظم استرجاع الإنترنت، عند الإشارة إلى النظم الخاصة باسترجاع المعلومات من الإنترنت، إلا أن هذا الكتاب يستعمل المصطلح الأخير، نظرا لأن الأول يبدو ضيقا

جدا في معناه. فالعنكبوتية، أولا، أحد التطبيقات لا أكثر، حتى وإن كانت التطبيق المسيطر في الإنترنت. كما أن المعلومات المتداولة من جانب التطبيقات الأخرى، كبروتوكول نقل الملفات (FTP) تدخل أيضا في نطاق نظم استرجاع الإنترنت. ومن ثم، فإن تسمية نظم استرجاع الإنترنت بمحركات البحث في العنكبوتية، يمكن أن تعني ضمنا أنها تتعامل مع المعلومات المتوافرة في العنكبوتية فقط، وتستبعد المعلومات الناشئة عن التطبيقات الأخرى للإنترنت. ثانيا، لبعض تطبيقات الإنترنت مقومات البحث الخاصة بها (مثل آرشي **Archie** بالنسبة لمعلومات بروتوكول نقل الملفات **FTP**)، وإن كان كثير من هذه التطبيقات قد توقف تدريجيا، أو بدأ استعمال العنكبوتية كمنصة. ثالثا، يمكن لأدوات استرجاع الإنترنت أن تكون معتمدة على الأدلة **directory-based** (مثل ياهو!) أو معتمدة على البحث **search-based** (مثل آلتافيستا **AltaVista**)، ومن ثم، فإنه قد يكون من غير الملائم الإشارة إلى نظم استرجاع الإنترنت المعتمدة على الأدلة، بوصفها محركات بحث في العنكبوتية؛ لأن البحث لم يكن يحظى بالدعم في تصميمها الأصلي. والمكون الخاص بالبحث المتاح الآن في كثير من نظم استرجاع الإنترنت المعتمدة على الأدلة، ناتج عن الاتحاد بين النوعين المختلفين من النظم، كما سنوضح فيما بعد.

وسوف نقدم في هذا الكتاب معلومات حول نظم استرجاع الإنترنت، أكثر من تلك التي نقدمها حول الأنواع الأخرى من نظم استرجاع المعلومات، لسببين؛ أولهما أن نظم استرجاع الإنترنت أحدث من نظم أخرى أعضاء في أسرة نظم استرجاع المعلومات. والسبب الثاني أن النظم الأخرى قد تمت دراستها بعمق، من جانب كثير من الباحثين على مر السنين (كنظم الخط المباشر من جانب هارتر **Harter, 1986**؛ ولارج، وتد، وهارتلي **Large, Teded, & Hartley, 1999**؛ وميدو، وبويس، وكرافت **Meadow, Boyce, & Kraft, 1999**؛ ووكر وجينز

Walker & Janes, 1999؛ ونظــم الأسـطوانات المكتنـزة **CD-ROM systems**، من جانــب شــودري **Chowdhury, 1999**، وراولي وسـلاك **Rowley & Slack, 1997**، والفهارس المتاحة على الخط المباشر **OPACs**، مـن جانـب بوليـو وبورجمـان **Beauliew & Borgman, 1996**؛ وهايلـدرث **Hildreth, 1985**). إلا أن نظـم استرجاع الإنترنت، لم تحظ بعد بالمعالجة على نحو منهجي شامل.

فئات نظم استرجاع الإنترنت:

تتوافر نظم استرجاع الإنترنت، على الرغم من حداثة عهدها، بأعداد ضـخمة وفئات متنوعة. ولم يكن من الممكن إحصاء المتاح من نظم استرجاع الإنترنت المتاحـة على الإنترنـت، نظـرا لأن هـذه الـنظم بـدأت تتزايـد في منتـصف تـسعينيات القـرن العشرين. إلا أن التقسيم الفئوي لنظم استرجاع الإنترنت، بناء على ما يلي من معايير، يقصد به إلقاء نظرة شاملة على أحدث أعضاء أسرة استرجاع المعلومات.

التصنيف وفقا لنهج الاسترجاع:

البحث والتصفح هما النهجان الرئيسان للاسترجاع، اللذان حظيا بالمعالجة في الفصل السادس. وعـلى الإنترنـت، تـسمى نظـم الاسـترجاع القائمـة عـلى البحـث بمحركات البحث (مثل آلتافيستا وجوجل) بينما تـسمى نظـم اسـترجاع المعلومـات القائمة على التصفح بالأدلة (مثل ياهو!). وغالبا ما تسمى الأدلـة أيـضا بالفهـارس. وتكفل محركات البحث للمستفيدين فرصة صياغة أو تكـوين استفـساراتهم الخاصـة بالبحث. وفي مقابل ذلك تنظم الأدلة، وتقدم موارد الشبكة تحت فئات منظمة هرميا. وبإمكان من يستخدمون نظم استرجاع الإنترنت القائمـة عـلى الأدلـة، العثـور عـلى المعلومات باتباع مسار محدد سلفا، أي هرم الفئات الذي يوضع للنظام.

وكما بينا في الفصل السادس، فإن لكل من البحث والتصفح كنهجين للاسترجاع، ما يكتنفه من قصور؛ فاستفسارات البحث، على سبيل المثال، من الصعب صياغتها، إلا أنها لا غنى عنها في نظم استرجاع الإنترنت القائمة على البحث. وعلى المستفيدين من الإنترنت صياغة استفساراتهم، نظرا لأن اختصاصيي المعلومات لا يعملون عادة كوسطاء في هذه البيئة. وعلى النحو نفسه، يمكن لعدد النتائج المسترجعة من الإنترنت، بالنسبة لمعظم الموضوعات، أن يتجاوز ببساطة الآلاف، إن لم يكن الملايين. إلا أن التصفح كنهج للاسترجاع، لا تتوافر له تقنية أو وسيلة لتضييق استفسار الاسترجاع. ومن ثم، فإن نظم استرجاع الإنترنت تحاول التغلب على مثل هذه المشكلات، بدعم كل من التصفح والبحث في الوقت نفسه. وتستخدم نظم استرجاع الإنترنت، القائمة على الأدلة، للسماح أو الترخيص لمحركات البحث بتوفير مقومات البحث على مواقعها. كذلك أيضا تعاقدت نظم استرجاع الإنترنت، القائمة على البحث، مع مرافق أو خدمات الأدلة لتوفير خيار التصفح في مواقعها. إلا أن الشراكات بين هذين الضربين من نظم استرجاع الإنترنت تتغير باستمرار، وأصبحت في نهاية الأمر لا معنى لها، عندما بدأ كبار الضالعين في هذا المجال، تطوير كل من محركات البحث والأدلة لنظمهم الخاصة في الوقت نفسه.

ويسفر الجمع بين التصفح والبحث في أي من المواقع، عن النوع الثالث من نظم استرجاع الإنترنت، أي النظم الهجين لاسترجاع الإنترنت. وبإمكان المستفيد في بيئة الهجين هذه البحث أو التصفح في الموقع نفسه دون الانتقال إلى آخر. وقد ترتب على ذلك ارتفاع مستوى فاعلية الاسترجاع.

التصنيف تبعا للتطبيق:

تطورت، منذ نشأة الإنترنت، الكثير من التطبيقات المختلفة: تلنت telnet(*)، وبروتوكول نقل الملفات FTP، والجوفر gopher، ونـدل المعلومـات واسـعة المـدى Wide area information servers (WAISs)، والعنكبوتيـة أو العنكبوتيـة العالمية World Wide Web، وهذا قليـل مـن كثـير. ويقـوم كـل واحـد مـن هـذه التطبيقات بإنجاز مهام معينة؛ فتلنت، على سبيل المثال، للارتباط من مسافة نائية، من نظام محلي، بحيث يمكن استغلال المقومـات أو المـوارد (كالعتـاد والبرمجيـات مـثلا) القائمة في الموقع النائي. أما بروتوكول نقل الملفات FTP فيـستخدم مـن أجـل نقـل الملفات، فيما بين النظم النائية والنظم المحلية بـسلاسة وبـسرعة عاليـة. والعنكبوتيـة Web هي أكبر مستودع للمعلومات على الإنترنت، نظرا لواجهة التعامل الخاصة بها، التي تراعي ظروف المستفيدين أو تتعاطف معهم، فضلا عن خاصية الوسائط الفائقة. وكان الجوفر، كتطبيق عاطل اليوم؛ إذ حلت محله العنكبوتية تمامـا، مـصمما لاختـزان واسترجاع المصادر على الإنترنت، اعتمادا على واجهة تعامل عبارة على قائمة خيـارات فائقة. ومن ناحية أخرى، كانت ندل المعلومات واسـعة المـدى أداة بحـث بـالكلمات المفتاحية في مصادر الإنترنت، متوافقة مع المواصفة المعيارية زد ٣٩.٥٠. وفي مرحلـة أو أخرى، كانت نظم الاسترجاع يتم بناؤها من أجل كل هذه التطبيقـات، فـيما عـدا ندل المعلومات واسعة المدى WAIS، نظـرا لأن هـذا التطبيـق الأخـير كـان يتكـون في حد ذاته من تقنية للاسترجاع. ويشتمل الجدول رقم ٨/ ١ عـلى نظـم الاسـترجاع التي نشأت من أجـل التطبيقـات الرئيـسة للإنترنـت، بنـاء عـلى نهجـي الاسـترجاع المختلفين(**).

(*) تلنت بروتوكول للشبكات، تطور عام ١٩٦٩، يستخدم في الإنترنـت أو الـشبكات المحليـة، لتـوفير مقومات التواصل التفاعلي القائم على النصوص، ثنائي الاتجاه، اعتمادا على منفذ افتراضي. (المترجم)

(**) للتعرف على مثل هذه التطبيقيات للإنترنت، الجاري منها والعاطل، يمكن الاطلاع على مقالة تصور واقع الإنترنت حتى منتـصف تـسعينيات القـرن العـشرين: حـشمت قاسـم. الإنترنـت ومستقبل خدمات المعلومات. دراسات عربية في المكتبات وعلم المعلومات، ع٢؛ ١٩٩٦. ص=

الجدول رقم ٨/ ١ تطبيقات الإنترنت وما يقابلها من نظم الاسترجاع

التطبيق	قائم عَلى البحث	قائم على الأدلة
تلنت	(متنوع)	هايتلنت
بروتوكول نقل الملفات	آرشي	(لا شيء)
العنكبوتية العالمية	آلتافيستا .. إلخ	ياهو!، ... إلخ
الجوفر	فيرونيكا، جاجهيد(١)	جواهر الجوفر(٢)
نـدل المعلومـات واسـعة المدى	كامن	(عن طريق الجوفر)

(١) Veronica, Jughead

(٢) Gopher Jewels

وكما يتضح من الجدول رقم ٨/ ١، فإن نظم استرجاع الإنترنت تشمل أكثر من مجرد محركات للبحث في العنكبوتية. وعلى الرغم من أن كثيرا من نظم الاسترجاع الواردة في الجدول رقم ٨/ ١ لم يعد لها وجود اليوم، فإنها كانت تنهض بدور متميز في استرجاع المعلومات من الإنترنت، وخصوصا قبل أن تتطور نظم استرجاع الإنترنت القائمة على العنكبوتية. ولمزيد من التحديد، فإن الهايتلنت(*) **Hytelnet** كانت تقدم قائمة تضم كثيرا من مصادر التلنت عن طريق واجهة تعامل نصوص فائقة. كذلك

=ص ٤٤-٨٨. وأعيد نشر هذه المقالة في: حشمت قاسم. الاتصال العلمي في البيئة الإلكترونية. القاهرة، دار غريب، ٢٠٠٥. ص ص ٤٠٧-٤٦٢.

(*) الهايتلنت محاولة مبكرة لتصميم واجهة تعامل عامة، أو بسيطة على الأقل، تتعامل مع مختلف مصادر المعلومات المعتمدة على تلنت، التي كانت متاحة قبل العنكبوتية العالمية. وكانت الهايتلنت تكفل لمن يتعاملون معها واجهة تعامل تصويرية **GUI**، تتيح فرصة تصفح دليل المصادر القائمة على تلنت، والوصول إليها على نحو معياري نسبيا. (المترجم)

كانت تكفل عمليات الربط الآلي للمواقع الـواردة في القائمـة. أمـا آرشي(*) **Archie** الذي قام بتطويره ألن إمتيج **Alan Emtage**، عندما كان طالبا في الدراسـات العليـا بجامعة ماجل **McGill** بكندا، فكان يقـوم دوريـا بمـسح مـضيفي بروتوكـول نقـل الملفات **FTP** مجهولي الاسم، وتجميع المعلومـات (كاسـم المـضيف، واسـم الـدليل، وحجم الملف على سبيل المثال) وذلك بالنسبة للملفات المتاحة في تلك النظم المضيفة. وكان بإمكان المستفيدين البحث في الملفات "القابلة للنقل بواسـطة بروتوكـول نقـل الملفات" باسم المضيف، واسم الدليل، واسم الملف، وذلك باسـتخدام آرشي. وكـان الجوفر يتمتـع بـشهرة كبـيرة قبـل أن تـصبح العنكبوتيـة هـي نجـم التطبيقـات عـلى الإنترنت. ولم يعد هناك الآن وجود على الإنترنت، لنظم الاسـترجاع التـي صـممت لمعلومات الجوفر، وهي فيرونيكا **Veronica**، وجاجهيد **Jughead**، وجواهر الجوفر **Gohper Jewels**؛ إذ توارى الجوفر تماما من الإنترنت. أما فترة حياة ندل المعلومات واسعة المدى **WAIS** فكانت قصيرة نسبيا، ويرجع ذلـك جزئيـا إلى واجهـة التعامـل الخاصة بها، التي كانـت تعتمـد عـلى لغـة الأوامـر، وكـذلك ظهـور نظـم اسـترجاع الإنترنت القائمة على العنكبوتية. إلا أن واجهة التعامل الموحدة، والمواصفة المعياريـة زد ٣٩.٥٠ المستقلة عن المنصة، التي تم تطبيقهـا في نـدل المعلومـات واسـعة المـدى، كانت تستخدم في كثير من تطبيقات الاسـترجاع الأخـرى (كالفهـارس المتاحـة عـلى الخط المباشر على سبيل المثال).

وقد تم استيعاب قطاعات من مهام الاسترجاع التي كان يتم إنجازها من قبـل بواسطة نظم لا تعتمد على العنكبوتية، كآرشي على سبيل المثال، في النظم القائمة عـلى العنكبوتية. ولم تعد نظم الاسـترجاع الأخـرى غـير القائمـة عـلى العنكبوتيـة (مثـل

(*) أداة لتكشيف أرشيفات بروتوكول نقـل الملفـات **FTP**، تكفـل للمتعـاملين معهـا العثـور عـلى ملفات بعينها. وتعد بمثابة أول محرك للبحث في الإنترنت. (المترجم)

فيرونيكا **Veronica**) متاحة الآن، نظرا لأن رسالتها الاسترجاعية قد تحققت. ومن الممكن أيضا استرجاع المعلومات، التي ترد من مثل هذه التطبيقات الخاصة بالإنترنت، مثل يوزنت **Usenet**(*) (بالنسبة للمناقشات التي تدور بين أطراف متعددة، حول موضوعات متعددة) والقوائم **lists** (الخاصة بالمناقشات التي تدور بين طرف واحد، وأطراف متعددة حول موضوعات منتقاة) قابلة للاسترجاع أيضا عن طريق نظم الاسترجاع القائمة على العنكبوتية العامة (مثل جوجل في **groups.google.com**، بالنسبة لمعلومات اليوزنت) أو نظم الاسترجاع المحددة (مثل **Tile.net at tile.net/lists** بالنسبة لمعلومات القوائم).

وقصارى القول، فإن نظم استرجاع الإنترنت في الوقت الراهن، تعتمد جميعها على العنكبوتية، أو تستخدم العنكبوتية كمنصة. وتغطي نظم الاسترجاع هذه المعلومات التي ترد من العنكبوتية، والتطبيقات الأخرى (مثل بروتوكول نقل الملفات **FTP**، ويوزنت **Usenet**) حتى وإن كانت العنكبوتية هي الآن أكبر مستودع للمعلومات على الإنترنت.

التصنيف وفقا للمحتوى:

من الممكن أيضا تقسيم نظم استرجاع الإنترنت إلى فئات، وفقا لمحتوى المعلومات التي تغطيها. فبعض هذه النظم يحرص على التغطية الشاملة (مثل جوجل) وذلك بتجميع المعلومات من مختلف المصادر والمجالات الموضوعية. وفي مقابل ذلك يمكن لنظم أخرى أن يتم بناؤها من أجل مجال واحد فقط بعينه. ومن أمثلة مثل هذه النظم بوابة المواد التربوية **Gateway for Educational Materials** (**www.thegateway.org**) بالنسبة للتربية.

(*) يوزنت، نظام لا مركزي عالمي للمناقشات في الإنترنت، نشأت فكرته عام ١٩٧٩، وتطور فعلا عام ١٩٨٠. (المترجم)

وقد تطورت بعض نظم استرجاع الإنترنت إلى مرافئ **portals**، بامتداد تغطية المحتوى إلى خارج نطاق المعلومات القابلة للبحث فيها. إلا أن البحث يظل هو البؤرة في المرافئ. والمرافئ الرأسية أو المتخصصة **vortals**، في مقابل ذلك، مرافئ تقتصر على قطاع رأسي (كالرعاية الصحية، أو التأمين، أو السيارات أو الصناعات الغذائية)؛ فالمجال أو القطاع أو السوق رأسي، لأنه يركز على نطاق ضيق نسبيا من السلع والخدمات، بينما السوق الأفقية هي تلك التي تهدف إلى إنتاج نطاق واسع من المنتجات والخدمات **(Tech Target, 2007)**. ويمكن القول بوجه عام، إنه من الممكن النظر إلى المرافئ بوصفها نظما عامة لاسترجاع الإنترنت. أما المرافئ الرأسية من ناحية أخرى، فنظم متخصصة لاسترجاع معلومات الإنترنت، موجهة نحو سوق رأسية. وكل من المرافئ والمرافئ الرأسية أو التخصصية، نظم هجين لاسترجاع معلومات الإنترنت، تدعم كلا من التصفح والبحث.

وبالإضافة إلى ذلك، هناك نظم الاسترجاع المكرسة لاسترجاع أنواع بعينها من المعلومات على الإنترنت. فأرقام الهاتف **Phone Number (www.phone number.com)**، على سبيل المثال، نظام يمكن أن يستخدم للحصول على العناوين البريدية وأرقام الهواتف. أما مابكويست **Map Qest (www.mapquest.com)**، فيكفل على سبيل المثال للمستفيدين طباعة عنوان ما، واسترجاع خريطة تفصيلية تفاعلية لذلك العنوان.

ويوضح هذا التصنيف الهرمي لنظم استرجاع الإنترنت، تنوع وقدرات نظم استرجاع المعلومات التي تطورت على مر السنين. وكما يبدو، فإن نظم استرجاع الإنترنت كفئة، قد تطورت وأصبحت أحدث، بل وأبرز الأعضاء في أسرة استرجاع المعلومات.

خصائص نظم استرجاع الإنترنت:

لم يتم تصميم نظم استرجاع الإنترنت، وخصوصا محركات وأدلة البحث في العنكبوتية، باستثناء وايز WAIS [ندل المعلومات واسعة المدى] بوصفها جزءا من جهود تطوير الإنترنت، وإنما نشأت هذه النظم بوصفها استدراكا، عندما كان المستفيدون من الإنترنت، وهم يواجهون كميات هائلة من المعلومات غير المجهزة، يفتقدون أي أداة لأغراض الاسترجاع؛ فآرشي Archie على سبيل المثال، قد تطور عندما كان المتعاملون مع بروتوكول نقل الملفات، يرغبون في معرفة ما إذا كانت أي مواقع مجهولة لبروتوكول نقل الملفات تستضيف المعلومات التي كانوا يحتاجون إليها. وقد صممت محركات وأدلة البحث في العنكبوتية، بعد أن واجه المستفيدون من العنكبوتية صعوبات متزايدة في العثور على ما كانوا يريدونه، بمجرد تتبع الروابط الفائقة الكامنة في مواقع العنكبوتية. ولو أن العنكبوتية برمتها عوملت بوصفها مكتبة، لكان من الممكن للموقف أن يكون قابلا للمقارنة بالحصول على مجموعات المكتبة، قبل إنشاء الفهرس الخاص بها. وتختلف نظم استرجاع الإنترنت، عن الأنواع الأخرى لنظم استرجاع المعلومات، في عدة أوجه أخرى إضافية نناقشها في الأقسام التالية.

التغطية والمعلومات المصدرية:

تمر أي معلومات، قبل أن تدخل في نظم الخط المباشر أو الفهارس المتاحة على الخط المباشر، حتما عبر التحرير والتحكيم أو أي إجراءات مراجعة مماثلة، لضمان جودتها. إلا أن هناك قليلا من تدابير ضبط الجودة في عملية إنتاج المعلومات على الإنترنت؛ فبإمكان أي إنسان وضع أي معلومات على الإنترنت، دون مراجعتها من منظور الجودة أو مدى الملاءمة. ويمكن للمعلومات من هذا النوع، أن تصبح المعلومات المصدرية بالنسبة لنظم استرجاع الإنترنت.

وتعتمد نظم استرجاع الإنترنت، في المقام الأول، على وسائل حصاد harvesting آلية، كالشواخص الآلية **robots**، أو العناكب **spiders**، أو الزواحف **crawlers**، لتجميع المعلومات من الإنترنت. كذلك تكفل هذه النظم للبشر تقديم المعلومات حول مواقعهم الخاصة، وإن كانت المعلومات التي تقدم ذاتيا، لا تشكل سوى نسبة ضئيلة من إجمالي المعلومات التي يتم تجميعها. وفيما عدا نظم استرجاع الإنترنت القائمة على الأدلة، فإنه يبدو من النادر أن نجد بشرا يختارون يدويا، معلومات من الإنترنت. وفي مقابل ذلك يتحمل اختصاصيو المعلومات مسئولية عملية تجميع البيانات في الأنواع الأخرى من بيئات الاسترجاع.

وعندما يتم إرسال أو إطلاق الشخوص الآلية أو العناكب، أو ما شابه ذلك، لتجميع البيانات، فإنها ليست جميعها مصممة لاستنساخ محتوى المواقع كاملة؛ فآرشي **Archie** على سبيل المثال، كان يستنسخ فقط معلومات الدليل الخاصة بمواقع بروتوكول نقل الملفات. وتلتقط بعض وسائل الحصاد الكلمات المئتين الأولى أو الأسطر العشرين الأولى، من كل موقع تقوم بزيارته. وتحصل وسائل أخرى على معلومات معينة فقط (كالعنوان، والرؤوس والروابط الفائقة على سبيل المثال) من المواقع. وبالإضافة إلى ذلك، فإن لمواقع العنكبوتية، بوصفها أكبر مستودع للمعلومات على الإنترنت، هياكلها الفائقة المتشابكة. وقليل من نظم استرجاع الإنترنت، ما يقصد أو تتوافر له المصادر لجمع المعلومات من كل وثيقة مرتبطة برابطة فائقة؛ فمن الممكن، على سبيل المثال، للبعض أن يقرر استنساخ الوثائق بناء على مستوى معين واحد للروابط الفائقة، تاركا المعلومات على المستويات الأخرى دون مساس. ومن ثم، فإنه بالنسبة لمعظم المواقع، لا تشتمل نظم استرجاع الإنترنت سوى على جزء فقط من معلوماتها الكاملة.

وبالإضافة إلى العاملين اللذين عرضنا لهما توا، فإن لما يسمى العنكبوتية الخفية أو غير المرئية **invisible** تأثيرها على تغطية نظم استرجاع الإنترنت. وتدل العنكبوتية الخفية على تلك المكونات التي لا يمكن الوصول إليها بواسطة النظم العامة أو المشتركة لاسترجاع الإنترنت، نظرا لصيغ الملفات (مثل الصور الناشئة عن استخدام الحاسب **Computer generated imagery CGI**) أو الهياكل **structures** (مثل مراصد البيانات المحملة على العنكبوتية). ولم تكن ملفات صيغ الوثائق متعددة المنصات **Portable Document Format (PDF)** على الإنترنت، تحظى بالتغطية من جانب أي نظام لاسترجاع الإنترنت، حتى غيَّر جوجل من ممارساته، في مطلع عام ٢٠٠١، لكي يستوعب هذه الملفات.

وبناء على دراسات أجراها لورانس وجايلز **Laurence and Giles (1998, 1999)**، فإن تغطية نظم استرجاع الإنترنت، بالنسبة للحجم المقدر للعنكبوتية القابل للتكشيف على الصعيد العام (أي ٨٠٠ مليون صفحة في فبراير عام ١٩٩٩) قد تراجعت على نحو ملحوظ، منذ ديسمبر عام ١٩٩٧؛ إذ لم يكن أي نظام يكشف أكثر من حوالي ١٦ بالمئة من الحجم المقدر للعنكبوتية القابلة للتكشيف على الصعيد العام. وفي عام ١٩٩٧ كانت نسبة التغطية حوالي ٣٣ بالمئة. وقد تم استخلاص النتائج النهائية بعد أن فحص الباحثون ستة من نظم استرجاع الإنترنت، في إحدى الدراسات، وأحد عشر نظاما في أخرى، اعتمادا على استفسارات حقيقية، أجراها العاملون في معهد بحوث إن إي سي **NEC**(*). وقد قدر آخرون (مثل **Smith,**

(*) إن إي سي، شركة يابانية عابرة للبحار، تقوم بتوريد خدمات ومنتجات تقنيات المعلومات والحلول الشبكية للمشروعات التجارية، ومتعهدي خدمات الاتصالات والأجهزة الحكومية، ومقرها الرئيس في ميناتو **Minato** أحد أحياء طوكيو باليابان. وكانت تعرف قبل عام ١٩٨٣ باسم شركة نيبون للكهرباء **Nippon Electric Company**. (المترجم)

2000) أن نظم استرجاع الإنترنت التي تتربع على القمة، لم تستطع تكشيف ما بين ٧٠ إلى ٧٥ بالمئة من صفحات العنكبوتية.

وأيا كانت التغطية الفعلية، فإن ما يبدو جليا، أن نظم استرجاع الإنترنت لا تستطيع سوى تجهيز نسبة متناقصة من معلومات الإنترنت، نظرا لأن الإنترنت تنمو أسيا. وتميز التغطية غير المكتملة للمعلومات، والممارسات غير الانتقائية للمعلومات المصدرية، نظم استرجاع الإنترنت، عما يقابلها في البيئات الأخرى لاسترجاع المعلومات.

تقنية التكشيف:

خوارزميات التكشيف الآلي، القائمة على إحصاء تكرار الكلمات والمعايير المماثلة، هي الغالبة في ممارسات التكشيف، في إنشاء نظم استرجاع الإنترنت. ويتطلب الأمر تدخلا بشريا محدودا في عملية التكشيف، حتى وإن كانت بعض النظم (ياهو! على سبيل المثال) تجنح إلى تقسيم المعلومات التي تشملها مراصد بياناتها، إلى فئات يدويا. ونادرا ما تستخدم اللغات المقيدة في تكشيف محتوى العنكبوتية؛ مراعاة لعوامل كالتكلفة، والفاعلية، وطبيعة المعلومات على الإنترنت.

وهناك بعض المشروعات (مثل إنتيوت **www.intute.ac.ukintute at**) التي تستعمل لغات مقيدة لتنظيم مصادر الإنترنت. ومن الممكن العثور على قائمة بالمشروعات المماثلة في سابيرستاكس **Cyberstacks**، وهو دليل افتراضي يعتمد على العنكبوتية، لكثير من المشروعات وثيقة الصلة بالمكتبات وعلم المعلومات (**McKiernan, 2001**). إلا أن هذه المشروعات، إذا أردنا الدقة والتحديد، إنما هي بمثابة فهارس متاحة على الخط المباشر، بالنسبة لمصادر الإنترنت، أكثر منها نظما لاسترجاع الإنترنت، نظرا لكثافة التدخل البشري فيها.

ويتم التحقق من حقول (كالمؤلف والعنوان وسنة النشر على سبيل المثال) وتكشيفها، عند بناء مراصد البيانات لنظم الخط المباشر ونظم الفهارس المتاحة على الخط المباشر. وفي مقابل ذلك لا تتكون نظم استرجاع الإنترنت من حقول بيانات بالمفهوم التقليدي، في مراصد بياناتها، خلاف الملفات الكشفية الخاصة بالكلمات المفتاحية والأماكن التي تقابلها في النظام، بلا حقول. ومن الممكن معاملة المعلومات المميزة **tagged**، مثل ...**<title> </title>** في لغة ترميز النصوص الفائقة (**HTML**)، أو ..**<price> </price>** في لغة الترميز القابلة للاتساع (**XML**) على أقصى تقدير، بوصفها أشباه حقول **quasi fields**، نظرا لأنها لا يتم تحديدها بطريقة موحدة في أثناء عملية التكشيف.

وقصارى القول، فإن تقنية التكشيف بالنسبة لنظم استرجاع الإنترنت، تختلف عن تلك الخاصة بالأنواع الأخرى لنظم استرجاع المعلومات، من حيث اتباع الأولى للنهج الآلي القائم على الكلمات المفتاحية، ولا تكشِّف مصادر الإنترنت وفقا للحقول.

مقومات البحث:

بوصفها أحد أنواع نظم استرجاع المعلومات، تكفل نظم استرجاع الإنترنت ، جميع مقومات البحث المتوافرة في بيئات الاسترجاع الأخرى تقريبا. إلا أن هناك بعض الجوانب الفريدة بالنسبة لنظم استرجاع الإنترنت نناقشها في هذا السياق.

فالبحث البوليني، المسلم به في غير نظم استرجاع الإنترنت، لم يكن يحظى فيما مضى بالدعم في كل نظام من نظم استرجاع الإنترنت. ومن ناحية أخرى كانت علامة زائد (+) تستعمل على نطاق واسع بديلا عن المعامل و**AND**، في السنوات المبكرة لنظم استرجاع الإنترنت، إلا أن هذا أمر مضلل، نظرا لأن علامة زائد تمثل الرمز

بالنسبة لمعامل أو **OR**، أو الحاصل **sum** المنطقي، كما بينا في القسم ٧/ ٢. وعلى الرغم من توقف هذا الشكل من الممارسات، لتصبح علامة زائد رمزًا للوزن بالنسبة للبحث الموزون، فإن استعمال علامة زائد بوصفها معامل و**AND**، يسبب الخلط أو التشويش بالنسبة للمستفيد النهائي، وخصوصا ذلك الذي لم يمارس البحث البوليني من قبل.

والبحث بالتقارب ضروري في استرجاع العبارات (إذا لم تكن معلومات مرصد البيانات مكشفة بالعبارات) فضلا عن تحديد المواقع النسبية لمصطلحات الاستفسار. وتتسم نظم الخط المباشر بالقوة، على وجه الخصوص، في دعم عمليات البحث بالتقارب. وفي مقابل ذلك لم يصبح البحث بالتقارب من الخصائص العامة لنظم استرجاع الإنترنت. ومن بين تلك النظم التي تدعم البحث بالتقارب على الإنترنت، يكفل القليل منها المدى الكامل لمعاملات التقارب، التي عادة ما تتوافر في نظم الخط المباشر والفهارس المتاحة على الخط المباشر.

وفيما يتعلق بالبتر، والبحث بالحقل، والبحث بمراعاة شكل الحرف الأول، فإن نظم استرجاع الإنترنت تدعم هذه التقنيات على نطاق محدود، وبقدرات محدودة. وإذا أخذنا البتر كمثال، فإن نظم الاسترجاع الأخرى يمكنها أن تكون محددة جدا، بتحديد عدد الأحرف التي يتم بترها، على سبيل المثال. أما معظم نظم استرجاع الإنترنت فيمكنها تحقيق البتر الأيمن غير المحدود فقط. ولا يزال البتر الآلي هو الخيار الافتراضي **default** في بعض الحالات، الأمر الذي يسفر عن المزيد من التشويش في نتائج البحث. وكما ذكرنا في القسم ٨/ ٤/ ٢/ ٢ فإنه ليس هناك سوى قليل من أشباه الحقول (مثل العنوان والمحدد الموحد لمكان المصدر **URL**) في نظم استرجاع الإنترنت، لأن مثل هذه النظم بلا حقول بالفعل. وليست هناك فرص كثيرة للمستفيد، لأن يحدد البحث بالحقل في نظم استرجاع الإنترنت.

ويظهر البحث الموزون، الذي قلما نجده في الأنواع الأخرى من نظم استرجاع المعلومات، ضمن رصيد مقومات البحث الخاصة بكثير من نظم استرجاع الإنترنت. وتستعمل علامة زائد (+) على نحو مطرد في نظم استرجاع الإنترنت بوصفها رمز الوزن الذي يضفي المزيد من الوزن على المصطلح المقصود. ويبدو أن البحث الموزون يتجاوز أو يتجاهل البحث البوليني على الإنترنت؛ ففي استفسار حول "الضرائب والغرامة أو العقوبة" **tax AND penalty**، على سبيل المثال، فإنه عند تحديد الوزن لـ **penalty**، يمكن للنتائج أن تشتمل على مصطلح **penalty** فقط، بينما يتم تجاهل **tax** كلية في عملية البحث.

والبحث المشوش **fuzzy**، على الرغم من أنه ليس بالخاصية المشتركة للأنواع الأخرى من نظم استرجاع المعلومات، ينفذ أو يطبق في بعض نظم استرجاع الإنترنت؛ فإذا كان المرء يقصد على سبيل المثال، البحث عن شارع روزفلت **Roosevelt Avenue** في خرائط ياهو!، فإنه يمكن لنظام استرجاع الإنترنت المتخصص في معلومات الخرائط، إذا ما حدث خطأ في هجاء عبارة شارع روزفلت **Roosevelt Avenue**، أن يظل قادرا على إرشاد من يقوم بالبحث إلى المكان الصحيح، إذا ما توافرت الأجزاء الأخرى من العنوان (كالرمز البريدي مثلا) بشكل صحيح. وخاصية جوجل "هل قصدت: **Did you mean:....**"، التي تظهر بعد وقوع الخطأ المطبعي، عند إدخال استفسار البحث، من تطبيقات البحث المشوش أيضا.

والبحث في مراصد بيانات متعددة ليس بجديد بالنسبة لنظم الخط المباشر. فعادة ما توضع مراصد البيانات الملائمة لتنفيذ مثل هذه الأنواع من عمليات البحث، أو تدار في نظام مضيف واحد؛ فنظام ديالوج **DIALOG** على سبيل المثال، قادر على دعم البحث في مراصد بيانات متعددة، وذلك بضم عدة مئات من مراصد البيانات المتشابهة من حيث البنيان. وفي حالة نظم استرجاع الإنترنت، يمكن لمصطلح البحث

في مراصد البيانات المتقاطعة **Cross-database searching** أن يكون مصطلحا أكثر ملاءمة في هذا السياق، نظرا لأن كل نظام قائم بذاته، مصمم على نحو مستقل، ومن ثم فإنها تختلف عن النظم الأخرى؛ فهذه النظم ليست مصممة كأقسام فرعية لكيان أكبر. وتسمى النظم التي يمكنها كفالة البحث في مراصد البيانات المتقاطعة، ما وراء محركات البحث **metasearch engines** (مثل دوجبايل **Dogpile**)، أو ما وراء نظم الاسترجاع **metaretrival systems**، لكي يكون الأمر متسقا مع المصطلحات المستعملة في هذا الكتاب. ويختلف عدد نظم الاسترجاع القائمة بذاتها، التي يمكن لما وراء محرك البحث تغطيتها، من محرك إلى آخر، وعادة ما يكون العدد في حدود تتراوح بين ستة نظم واثني عشر نظاما. وعادة ما يعمل البحث بالكلمات المفتاحية المفردة على أحسن وجه، في مراصد البيانات المتقاطعة، نظرا لأن هناك أوجه اختلاف أكثر من مظاهر التشابه في نَظْم **syntax** مصطلحات البحث ومعانيها، فيما بين نظم الاسترجاع القائمة بذاتها تلك. ويستبعد بعض ما وراء نظم الاسترجاع في الإنترنت المكررات قبل تقديم النتائج للمستفيد.

والبحث القائم على المفهوم أو المعنى، لا على مضاهاة الكلمات المفتاحية، هو المفضل على وجه الخصوص على الإنترنت، نظرا لأنه حاسم في الارتفاع بمستوى أداء الاسترجاع. كما أن اللغات المقيدة لا تستعمل عادة في نظم استرجاع الإنترنت. وقد جرت عدة محاولات في هذا الصدد؛ فمن الممكن على سبيل المثال، لنظام استرجاع المعلومات أن يولِّد آليا قائمة بالمفاهيم، بناء على مصطلحات الاستفسارات التي يتم إدخالها، ويمكن للمستفيد أن يختار منها؛ فإذا كان الاستفسار على سبيل المثال، "احجز رحلة" **book a flight**، فإن النظام يقدم قائمة خيارات منسدلة، تشتمل على مفاهيم "كالسفر جوا" **air travel**، و"رحلة" **trip**،و"مسار أو خط سير" **trajectory**. وحينئذ يختار المستفيد مفهوما مقابلا (مثل "السفر جوا" **air travel**)

لكي يواصل عملية البحث الفعلي. وهناك مثال آخر يتصل ببحث المفاهيم، يكفل للمستفيد القدرة على إجراء عمليات البحث دلاليا. وفي هذا التطبيق، لا يمكن لغياب كلمة ما من الوثيقة (''مترو الأنفاق'' **subway** مثلا) أن تستبعد احتمال أن تكون الوثيقة صالحة موضوعيا أو مفاهيميا للاستفسار حول ''النقل العام'' **public transportation** على سبيل المثال. ويمكن القول على وجه التحديد، أن مثل هذه التطبيقات لا يمكن النظر إليها إلا بوصفها مجرد أشباه بحث مفاهيمي، نظرا لأن تكشيفها لا يبدو قائما على المفاهيم، وهذا من الشروط المسبقة للبحث المفاهيمي. وبالأحرى فإنه يتم رسم خريطة للمفاهيم بالكلمات المفتاحية في أثناء عمليات البحث، باستعمال أدوات معينة (قائمة بالمترادفات على سبيل المثال). وفي الأنواع الأخرى من نظم استرجاع المعلومات، يكاد استعمال اللغات المقيدة يكون هو القاعدة، حيث يحظى البحث بالمفاهيم بالدعم بوجه عام. وفضلا عن ذلك فإنه عادة ما يصبح التعامل مع نظم استرجاع الإنترنت، التي تدعم البحث، مقابل رسوم، بمجرد أن تكمل اختبارها النهائي على الإنترنت. ويمكن أن تكون العنكبوتية الدلالية، التي نتناولها في القسم ١٢/ ٣، قادرة على أن تكفل البحث المفاهيمي بلا قيد ولا شرط، عندما يتم إنشاؤها.

وبحث الأنداد **peer-to-peer (P2P)**، إحدى الخواص الجديدة في حقل استرجاع المعلومات، ويرتبط ارتباطا وثيقا بنهج التنقية التضامنية أو التعاونية **collaborative filtering**، المتبع في التسويق التجاري. إذ يتم فحص الخلفيات التاريخية، والإشارات المرجعية الخاصة بالمستفيدين السابقين، عند تلقي استفسار جديد، من أجل اكتشاف المعلومات المرتبطة بالموضوع. وعندما ينضم المرء إلى شبكة الأنداد **P2P**، فإنه يتم بحث استفساره مقابل المواقع التي زارها وأشار إليها من ينتمون إلى الشبكة المبدئية نفسها. ومن الممكن تكرار هذه العملية بتوسعة مجال

البحث، ليشمل شبكات الأنداد الإضافية التي ينتمي إليها هؤلاء. وقد تناول فستا **(2001) Festa**، اعتمادا على باندانجو **Pandango**، وهو مشروع قامت بتنفيذه آي فايف الرقمية **i5 digital**، كمثال:

> يمكن ... لباندانجو أن يقر الارتباط بالموضوع بفحص شبكة مشعة **radiating** من "من يسجلون الإشارات المرجعية **referrers**". وبمجرد أن يقوم شخص ما بتحميل باندانجو **Pandango**، وينضم إلى شبكة أنداد ما، فإنه يمكن لبحث ذلك الشخص بالكلمات المفتاحية، أن يفحص الخلفيات التاريخية للعنكبوتية، والإشارات المرجعية لشبكة مبدئية لمئة ممن يسجلون الإشارات. وحينئذ يمكن للتطبيق أن يبحث الخلفيات التاريخية لأولئك المئة على العنكبوتية، وجمع عشرة آلاف ممن يسجلون الإشارات، ويعيد ذلك مرة أخرى، بحيث يمكن للاستفسار المرور على صفحات العنكبوتية التي تمت زيارتها والإشارة إليها من جانب مليون شخص. **(P.1)**

وبحث الأنداد لا مركزي، ويفيد من معلومات البحث الناشئة عن مستفيدين سابقين، بدلا من البحث في مرصد بيانات مركزي. ويمكن لقضايا معينة (كالخصوصية والأمن على سبيل المثال) أن تثار حتما نتيجة لبحث الأنداد على الإنترنت. إلا أن هذه القضايا لا تدخل في مجال اهتمام هذا الكتاب. كذلك يسمى بحث الأنداد بالرابط **bonding** في بعض الوثائق (مثل **Schwartz, 2000**). ومن ناحية أخرى يمكن لحركة الجيل الثاني للعنكبوتية **Web 2.0** أن تحل محل بحث الأنداد بخواص إضافية. ونستكشف تأثير تقنيات الجيل الثاني للعنكبوتية **Web 2.0**، على نظم استرجاع المعلومات في القسم رقم ٨/ ٥.

تقنيات الترتيب الطبقي:

دخلت طرق الترتيب الطبقي **ranking** في استرجاع معلومات الإنترنت، كجهد يهدف لمساعدة المستفيد في العثور على الإبر في كومة التبن، على نحو أكثر فاعلية. وتشمل الخوارزميات التي عادة ما تطبق للترتيب الطبقي، لنتائج عمليات البحث في جميع بيئات استرجاع المعلومات، كما تناولناها في القسم ٥/ ١/ ٢/ ٢، تردد المصطلحات، وتجاور المصطلحات، ومواقع المصطلحات، والتردد العكسي للوثائق. وبينما ترتب هذه الخوارزميات نتائج عمليات البحث وفقا للخصائص الجوهرية للوثائق (تردد المصطلحات على سبيل المثال)، هناك طرق ترتيب طبقي جديدة، تركز على الخصائص الجوهرية الناشئة عن البشر عندما يصلوا إلى وثائق العنكبوتية المستهدفة، أو عندما ينشئون مواقعهم الخاصة.

وتستند أول تقنية جديدة للترتيب الطبقي، التي كان لمؤسسي جوجل، سيرجي برين **Sergey Brin** ولورنس بيج **Lawrence Page**، فضل الريادة فيها، على الروابط الراجعة **backlinks**، أي الروابط التي تشير إلى الموقع المزمع استرجاعه **(Brin & Page, 1998)**. وتفترض هذه التقنية المسماة بترتيب بيج الطبقي **PageRank**، أنه كلما زاد عدد مرات الإشارة إلى الموقع من جانب مواقع أخرى، أصبح أكثر أهمية. وتبدو هذه التقنية مشابهة إلى حد بعيد لتحليل الاستشهادات المرجعية، إحدى الطرق الراسخة التي تستخدم ضمن أشياء أخرى، لتقييم جودة الوثائق التخصصية. وباستعارة بعض المصطلحات من تحليل الاستشهادات المرجعية، يمكننا تسمية الموقع الذي يشار إليه بالموقع المستشهد به، بينما يصبح الموقع الذي يشير إلى الموقع المستشهد به الموقع المصدري، الذي يرد به الاستشهاد **citing site**. وبعبارة أخرى، طور جوجل طريقة لتحليل الروابط التي تشير إلى المواقع المستشهد بها، وترتيب هذه المواقع طبقيا. واسم ترتيب بيج الطبقي **PageRank** خوارزمية تلاعب بالكلمات؛ لأنه سمي باسم أحد واضعيه، وهو لورنس بيج، ويدل على المستهدف

بالترتيب الطبقي وهو صفحات العنكبوتية. وتوضع المحددات الأخرى كالعنوان، وأشكال الأحرف، وتقارب المصطلحات، وأهمية المواقع المصدرية التي ترد بها الاستشهادات، في الحسبان أيضا، في حساب الرتبة **(Brin & Page, 1998)**.

وقد تطورت خوارزمية الترتيب الطبقي الجديدة الثانية، التي تعرف بنهج الرواج أو الشيوع **popularity approach**، في دايركت هيت(*) **Direct Hit**، التي أصبحت فيما بعد جزءا من تيؤما(**) **Teoma**. ثم دخل تيؤما فيما بعد في حوزة آسك دوت كوم(***) **Ask.com**، وهو نظام آخر لاسترجاع الإنترنت. وترتب هذه الخوارزمية المواقع طبقيا بناء على عدد مرات زيارتها فعلا من جانب المستفيدين (ومن هنا جاءت كلمة التسديدة **hit** في **Direct Hit**). فإذا نظرنا إلى آلتافيستا **AltaVista** وما يناظره من محركات البحث، بوصفها نظما لاسترجاع المعلومات، يتحكم فيها المؤلفون (أي أن صلاحية نتائج عمليات البحث أو اتصالها بالموضوع، تتقرر بناء على مدى مضاهاة الكلمات المفتاحية لمحتوى الوثيقة)، وإذا ما أسمينا ياهو! وما يناظره من أدلة نظما لاسترجاع المعلومات يتحكم فيها المحررون (أي أن المحررين يعثرون على المواقع ويفهرسونها بفحصها كلا على حدة)، إذن فالتسديدة المباشرة **Direct Hit** تمثل نوعا ثالثا من طرق البحث، وهو نظم استرجاع المعلومات التي يتحكم فيها المستفيد، ويتوقف فيها الترتيب الطبقي لنتائج عمليات البحث على الخيارات التي يقرها المستفيدون أنفسهم **(Frauenfelder, 1998)**. وكلما زاد عدد زيارات الموقع من جانب آخرين، يزداد الموقع رواجا، ويتحسن وضعه الطبقي. إلا أن هذه الخوارزمية على وجه الخصوص، لم تطبق على نطاق واسع.

(*) كانت إحدى شركات محركات البحث، ومقرها بوسطون. (المترجم)

(**) تيؤما محرك بحث في الإنترنت، تطور عام ٢٠٠٠. (المترجم)

(***) آسك دوت كوم، محرك بحث في العنكبوتية، تطور عام ١٩٩٦، ويركز على الإجابة عن الأسئلة. (المترجم)

وتنطوي خوارزمية الترتيب الطبقي الثالثة على خطوتين؛ أولاهما ترتيب المواقع طبقيا، باتباع طرق الترتيب الطبقي التقليدية كتردد المصطلحات، وتقارب المصطلحات، من أجل تحليل الروابط الفائقة، المتضمنة في الموقع والنص المحيط بهذه الروابط. والخطوة الثانية هي إعادة تدقيق هذه الروابط الفائقة باتباع طريقة الروابط الراجعة التي يستخدمها جوجل. وقد تطورت طريقة الترتيب الطبقي المسماة هيتس **HITS** (اختيار موضوع النص الفائق الناجم **hypertext induced topic selection**)، على يدي جون كلاينبرج(1999) **Jon Kleinberge**، عندما كان يعمل بمشروع آي بي إم **IBM** كليفر **CLEVER**. ويهدف كليفر للتحقق من النوعين التاليين من مواقع العنكبوتية، وهما: (١) الحجج **authorities**، أي أفضل مصادر المعلومات حول موضوع معين، و(٢) الحواشي **hubs**، أي مجموعات روابط تلك المواقع. ويؤدي تكرار هاتين الخطوتين إلى العثور على نتائج البحث وتنقيتها، تلك النتائج التي تقدم في النهاية بوصفها صفحات الروابط الصالحة أو المتصلة بالموضوع، مقسمة إلى حواشٍ وحجج. ويمكن للحواشي والحجج أن تحتل مراتب أعلى من المواقع العادية. وكليفر رسميا، أحد مشروعات بحوث آي بي إم. ولم يقدم حتى الآن سوى عمليات بحث وتقارير بحوث للعرض، أما خوارزمية هيتس **HITS** فترتبط ارتباطا تكامليا على نحو متزايد، بكثير من طرق الترتيب الطبقي في نظم استرجاع الإنترنت.

وطريقة الروابط الراجعة، التي استخدمت أولا من جانب جوجل، ثم فيما بعد من جانب كثير من نظم استرجاع الإنترنت الأخرى، ليست جديدة في بيئة الاسترجاع على الخط المباشر؛ فقد صمم بمعهد المعلومات العلمية(*) **Institute for**

(*) أصبح الآن في حوزة شركة **Reuters - Thopson**، واتسع نشاطه في مراصد بيانات الاستشهادات المرجعية، متمثلا في عنكبوتية المعرفة **Web of Knowledge**، وعنكبوتية العلوم **Web of Science**. (المترجم)

Scientific Information (ISI) سلسلة من مراصد البيانات (مثل كشاف الاستشهاد المرجعي في العلوم **Science Citation Index**) منذ ستينيات القرن العشرين، وفقا لمبدأ الاستشهاد المرجعي(*)، إلا أن تكرار الاستشهاد المرجعي لم يكن يستخدم كمعيار للترتيب الطبقي لنتائج عمليات البحث في مراصد بيانات الاستشهاد المرجعي، حتى تم إدخال الأمر "رتب طبقيا" **RANK** من جانب ديالوج **DIALOG** في عام ١٩٩٣. أما طريقة التسديدات المباشرة **Direct Hits** للرواج، وخوارزمية تسديدات كلاينبرج **Kleinberg's HITS**، من ناحية أخرى، فليس هناك ما يقابلهما في بيئات الاسترجاع الأخرى؛ حيث لا تزال طرق الترتيب الطبقي غير متوافرة على نطاق واسع.

وبينما يمكن لطرق الترتيب الطبقي التقليدية أن تسفر عن مشكلات، كحشو الكلمات **Stuffing** (أي تكرار بعض الكلمات المفتاحية عمداً للحصول على مرتبة أفضل)، أو لا تؤدي أي معنى يذكر، في مواقف مثل الدفع في مقابل المكانة **pay for placement** (أي ترتيب نتائج عمليات البحث طبقيا بناء على كم ما يسدد من رسوم) فإن التقنيات الثلاث التي تناولناها هنا، تفتح بعدا جديدا للترتيب الطبقي لمواقع العنكبوتية وتقييمها. إلا أنها يمكن أن تسفر عن المزيد من مظاهر التحيز في عملية الترتيب الطبقي؛ لأنه يمكن للمواقع الجديدة غير المرتبطة أن تواجه وقتا متزايد الصعوبة حتى تصبح بارزة في نظم الاسترجاع تلك **(Lawrence & Giles, 1999)**.

(*) لمزيد من المعلومات حول كشافات الاستشهاد المرجعي، راجع: حشمت قاسم. كشافات الاستشهاد المرجعي وإمكاناتها الاسترجاعية. المجلة العربية للمعلومات، مج٢، ع٤؛ يونيو ١٩٨٠. وأعيد النشر في: حشمت قاسم. دراسات في علم المعلومات. ط٢. القاهرة، دار غريب، ١٩٩٥، ص ص ٣٣٥-٣٥٣. (المترجم)

وليست خوارزميات الترتيب الطبقي القائمة على الروابط، بمنأى عن إساءة الاستخدام والانتقادات. وتشمل التدابير الرامية إلى استغلال نهج الروابط الراجعة مزارع الروابط **link farms**، وصفحات المداخل **doorway pages**. ومزارع الروابط مجموعات من الروابط التي تشير إلى أي صفحة أخرى في موقع معين، من أجل تعزيز إحصاء الروابط الراجعة لذلك الموقع، بينما صفحات المداخل عبارة عن صفحات عنكبوتية تتكون كاملة من روابط لتحقيق الهدف نفسه. ويمكن أن تستخدم طريقة التسديدات **HITS** للتصدي جزئيا للمشكلات الناشئة عن مزارع الروابط وصفحات المداخل.

وبسبب تلك الإجراءات التعسفية أيضا، يتم استعمال أكثر من مئتي إشارة (أي عوامل) في الترتيب الطبقي لنتائج عمليات البحث الخاصة باستفسار معين، من جانب جوجل **Google (2008b)**. وقد لا تستعمل نظم استرجاع الإنترنت الأخرى، ذلك العدد الكبير من المعايير كما يفعل جوجل، في الترتيب الطبقي لنتائج البحث. إلا أنه بعد التحقق من هذه الأنواع من الإجراءات التعسفية، لم يعد بإمكان أي نظام لاسترجاع معلومات الإنترنت، الاعتماد على أي طريقة بعينها قائمة على الروابط لأغراض الترتيب الطبقي.

وانتقادات الترتيب الطبقي القائم على الروابط، متعددة الجوانب بوجه عام؛ إذ يتم أولا احتساب كل رابطة بوصفها صوتا واحدا للصفحة المرتبطة، دون التمييز بين الروابط. ثانيا، لا تبذل أي جهود للتحقق من مكان صفحة العنكبوتية التي تشير إليها الرابطة؛ فالرابطة بعنوان الصفحة على سبيل المثال، تعامل على النحو نفسه، الذي تعامل به الرابطة بعنوان البريد الإلكتروني الخاص بمسئول العنكبوتية **webmaster**، الذي يمكن أن يكون مسئولا عن تطبيق لغة ترميز النصوص الفائقة بالنسبة للصفحة فقط، وليس عن محتواها الفكري على الإطلاق. ثالثا، يتم احتساب الروابط الذاتية

self-links، والروابط المكررة، والروابط من مراصد البيانات المترابطة فيما بينها، والمواقع التي تكرر مواقع قائمة **mirror sites** فعلا، يتم احتساب كل هذه جميعا بوصفها مساوية للروابط العادية (**Lewandowski, 2005**). وتردد مثل هذه الانتقادات، وإلى حد بعيد، صدى تلك الانتقادات التي توجه للمؤشرات القائمة على الاستشهاد المرجعي، في القياسات الوراقية، وقياسات النشاط العلمي **scientometrics**. فدوافع الربط، شأنها في ذلك شأن أسباب الاستشهاد المرجعي(*)، متعددة، ومعقدة، ومن الصعب التحقق منها نظرا لطبيعة أسلوب

الربط (**Chu, 2005**). وفضلا عن ذلك بيَّن بهارات وهنسنجر **Bharat and Hensinger (1998)** أن الحواشي والحجج المرتبطة ارتباطا وثيقا، التي يتم اكتشافها بطريقة التسديدات **HITS**، قد لا تكون حول الموضوع الأصلي.

تعديل البحث:

يحظى تعديل استفسارات أو صيغ البحث، في نظم الاسترجاع على الخط المباشر بالدعم المناسب؛ فبإمكان المستفيد توسعة أو تضييق عملية البحث بمختلف الطرق المتاحة في النظام (كالبتر أو البحث بالحقل على سبيل المثال). وبالمقارنة بنظم الخط المباشر، تبدو الفهارس المتاحة على الخط المباشر، أقل مرونة في تعديل صيغ البحث، على الرغم من تحسن الوضع في أجيالها التالية. يضاف إلى ذلك أن الحاجة إلى تعديل عمليات البحث في الفهارس المتاحة على الخط المباشر محدودة نسبيا؛ نظرا لأن

(*) لمزيد من المعلومات حول علاقة الوثائق التي ترد بها الاستشهادات **source documents**، والوثائق المستشهد بها **cited documents**، راجع:

حشمت قاسم. تحليل الاستشهادات المرجعية وتطور القياسات الوراقية. المجلة العربية للمعلومات، مج٣، ع٥؛ ديسمبر ١٩٨٠. ص ص ١١-٤٧. وأعيد نشر هذه المقالة في: حشمت قاسم. دراسات في علم المعلومات. ط٢. القاهرة، دار غريب، ١٩٩٥، ص ص ١١٧-١٤٧.

البحث عن أوعية معروفة يشكل نسبة لا يستهان بها من إجمالي عمليات البحث في الفهارس المتاحة على الخط المباشر.

وفي مقابل ذلك، فإن عمليات البحث في الإنترنت عادة ما تتطلب تعديلا، نظرا للكميات الضخمة من النتائج التي تسفر عنها عمليات البحث، وما يكتنف السماح للمستفيد النهائي بصياغة استفسار محكم من صعوبات؛ لأنه أولا لا يبدو من السهل بالنسبة للمستفيدين تعديل عمليات بحثهم؛ فليست هناك أولا فئات يتم إنشاؤها واختزانها لكل عملية بحث، بينما يعد ذلك من الخصائص المألوفة في نظم الخط المباشر. ثانيا، لا تتوافر سوى طرق للتعديل المحدود، بالنسبة لإحكام عمليات البحث؛ فجوجل، على سبيل المثال، يكفل للمستفيدين إجراء المزيد من عمليات البحث، في حدود النتائج التي يتم استرجاعها، بمجرد أن يتم الاسترجاع، وذلك عن طريق اللجوء إلى خيار "البحث في حدود النتائج". وذلك في الأساس هو ما يمكن للمستفيد الحصول عليه في نظم استرجاع الإنترنت الراهنة، فيما يتعلق بتعديل البحث. ولا تزال مرونة وتنوع خيارات تعديل عمليات البحث، التي تحظى بالدعم في بيئات الاسترجاع الأخرى، مفتقدة على الإنترنت.

ولما كانت اللغات المقيدة لا تطبق بوجه عام في تجهيز معلومات الشبكات، فإن بعض النظم (مثل جيجا بلاست **Gigablast**)(*)، تحاول بناء على الاستفسار الذي يدخله المستفيد، اقتراح قائمة بالمصطلحات من أجل تعديل عمليات البحث. فعندما يتم على سبيل المثال، إدخال استفسار عن الزنبق أو الخزامى **tulips**، يقدم جيجا بلاست المصطلحات التالية للمزيد من إحكام عملية البحث: **Holland, plant, spring, tulip bulbs**، وأكثر من ذلك. وبإمكان المستفيد اختيار أي من المصطلحات الواردة في القائمة، لاستعماله في تعديل البحث. إلا أنه على الرغم من

(*) محرك بحث في الإنترنت، منفتح المصدر. (المترجم)

توافر مثل هذه الخواص التي تطبق في بعض النظم، لا يزال من الصعب بمكان على المستفيدين من الإنترنت، تضييق عمليات البحث إلى المستوى المرغوب فيه. إلا أنه هناك كل من التلقيم المرتد الخاص بالصلاحية **relevance feedback**، والطريقة الآلية لتوسعة عمليات البحث، في نظم مثل جوجل. وبإمكان المستفيد استدعاء التلقيم المرتد الخاص بالصلاحية، وذلك ببساطة بالنقر على الروابط الفائقة المسماة بـ ''الصفحات المثيلة'' أو ''المزيد مثل هذا'' على سبيل المثال. وهناك في الوقت نفسه مجال للمخاوف بشأن مدى جدوى هذه الطريقة، إذا ما كان المستفيد يشعر فعلا بالغرق في عدد نتائج عملية البحث التي استرجعها النظام.

واجهة التعامل:

عادة ما تكون لغة الأوامر والاختيار من قوائم الخيارات، هما طريقتا واجهة التعامل السائدتان في نسبة كبيرة من نظم الخط المباشر، والفهارس المتاحة على الخط المباشر. ويفسر ذلك على الأقل لما كان لبعض نظم استرجاع المعلومات، بالنسبة لغير تطبيقات العنكبوتية (مثل آرشي **Archie** وفيرونيكا **Veronica**) واجهات تعامل من أنواع مماثلة. وعندما تصبح العنكبوتية تطبيقا رئيسا للإنترنت، ومنصة مألوفة للتطبيقات الأخرى (كالبريد الإلكتروني مثلا) تميل جميع نظم استرجاع الإنترنت، لأن تكون لها واجهات تعامل مع العنكبوتية، وهي واجهات تصويرية فائقة البناء أو التنظيم. ويسهم هذا التوحيد في واجهات التعامل في تعاطف نظم استرجاع الإنترنت ككل مع المستفيد، وهي خاصية ليست جوهرية بالنسبة للأنواع الأخرى من نظم استرجاع المعلومات.

وقصارى القول، فإن لنظم استرجاع الإنترنت، خواصها الفريدة في التغطية، والمعلومات المصدرية، والتكشيف، وإمكانيات البحث، وطرق الترتيب الطبقي، وتعديل البحث، وواجهات التعامل، وذلك عندما تقارن بنظم الخط المباشر،

والفهارس المتاحة على الخط المباشر. ويمكن للبحث عن المعلومات متعددة اللغات، متعددة الوسائط أن يميز أيضا نظم استرجاع الإنترنت عن الأنواع الأخرى من نظم استرجاع المعلومات. ونغطي ذلك الموضوع في الفصل التاسع "استرجاع المعلومات الفريدة في المحتوى أو في الشكل".

أجيال نظم استرجاع الإنترنت:

لقد تطورت نظم استرجاع الإنترنت، لتصبح عضوا جديدا في أسرة استرجاع المعلومات، ولتصبح لا غنى عنها بالنسبة لكثير من المستفيدين **(Fallows, 2008; Fallows, Rainie & Mudd, 2004)**. وقد أُدخلت الكثير من التعديلات والتحسينات على هذه النظم منذ بدايتها. وعلى الرغم من حداثة عهدها في التطور، مقارنة بما يقابلها، يمكن تقسيم نظم استرجاع الإنترنت إلى ثلاثة أجيال مختلفة، بناء على الخواص الواردة في الجدول رقم ٨/ ٢.

ويدل الجيل الأول لنظم استرجاع الإنترنت، على مثل هذه النظم الرائدة مثل ياهو! وزواحف العنكبوتية **Web Crawler**. فقد كانت المعلومات النصية هي النوع الرئيس للمعلومات التي كانت تجمعها تلك النظم. ولم يكن تجهيز المعلومات متعددة الوسائط يحظى باهتمام يذكر، عندما تم بناء هذا الجيل من نظم استرجاع الإنترنت، إلى جانب عرض الصور الكامنة فعلا في الصفحات أو المواقع، جنبا إلى جنب مع المعلومات النصية التي يتم استرجاعها. وكانت الأدلة ومحركات البحث متميزة عن بعضها البعض، وكان كل نظام يكفل طريقة واحدة للاسترجاع، سواء كانت التصفح أو البحث. وكانت هذه النظم تكشف بالكلمات المفتاحية، وتدعم كجيل أول، إمكانيات البحث الأساس (كالبحث البوليني، والبتر، والبحث بالتقارب على سبيل المثال). ولم يكن تعديل البحث ممكنا في هذا الجيل من النظم. ولم تكن مخرجات عمليات البحث ترتب طبقيا، إلا أنه كانت تتاح للمستفيدين فرصة اختيار شكل العرض (كالعنوان فقط، أو العنوان مع ملخص موجز على سبيل المثال) وكان بإمكان النظام عرض عدد النتائج (١٠ أو ٢٠ أو ٣٠ على سبيل المثال) دون تجديد الشاشة.

الجدول رقم ٨/ ٢ أجيال نظم استرجاع الإنترنت

الجيل / الخاصية	الجيل الأول	الجيل الثاني	الجيل الثالث
المعلومات المغطاة	نصوص في المقام الأول	نصية أو غير نصية	متعددة الوسائط
طريقة الاسترجاع	التصفح والبحث مستقلين	التصفح والبحث مجتمعين	التصفح والبحث متكاملين
التكشيف	الكلمات المفتاحية	الكلمات المفتاحية مع بيان العلاقة بين المفاهيم	الكلمات المفتاحية والمفاهيم
إمكانيات البحث	الأساس	الأساس والمتقدم	الارتفاع بمستوى التحقيق
تعديل البحث	لا يوجد	دعم محدود	المزيد من مقومات التضييق
المخرجات	غير مرتبة طبقيا	مرتبة طبقيا	مهيأة وفقا لرغبات الشخص المستفيد

وكان الجيل الثاني لنظم استرجاع الإنترنت أكبر على نحو ملحوظ من الجيل الأول. وكانت بعض النظم تمثل تحسينات على طبعاتها المبدئية (مثل ياهو!) بينما كانت نظم أخرى تمثل إضافات جديدة (مثل جوجل). ولم يكن عدد لا يستهان به من نظم الجيل الثاني يكفل خدمات استرجاع معلومات نصية فحسب، وإنما كانت تساعد

أيضا في العثور على ملفات الصور ومقاطع الفيديو والتسجيلات السمعية والموسيقية إم بي ثري **MP3**. إلا أن عملية الاسترجاع كانت محدودة، وتقتصر على مراصد البيانات أو الملفات التي تشتمل على نوع بعينه من معلومات الوسائط المتعددة. وبعبارة أخرى فإن المعلومات النصية ومعلومات الوسائط المتعددة كانت تجهز بمعزل عن بعضها البعض، في هذا الجيل من نظم استرجاع الإنترنت. وعلى الرغم من إدارتها بواسطة شركات مختلفة في بعض الحالات، كان نهجا التصفح والبحث يتوافران في الموقع ذاته، بحيث يمكن للمستفيد أن يختار من بين نهجي الاسترجاع دون الانتقال إلى موقع آخر. وكان التكشيف لا يزال يعتمد على الكلمات المفتاحية، إلا أن بعض الجهود كانت تبذل لرسم خرائط المفاهيم في شكل كلمات مفتاحية، باستخدام وسائل للبحث عن المفاهيم (راجع القسم ٨/ ٤/ ٢/ ٣). وبينما ظل البحث الأساس يحظى بالدعم من جانب الجيل الثاني لنظم استرجاع الإنترنت، فقد تم إدخال إمكانيات بحث متقدمة (كالبحث الموزون على سبيل المثال). وتكفل معظم نظم استرجاع الإنترنت نوعين من إمكانيات البحث؛ (١) البحث البسيط، الذي يتضمن عمليات بحث خالية من الرموز أو الإشارات (كالمعاملات على سبيل المثال)، و(٢) البحث المتقدم الذي يقبل الرموز والإشارات. وتستخدم نماذج **forms** البحث عادة لعمليات البحث المتقدم. وكان من الممكن توسعة عمليات البحث أو تضييقها، بعدد محدود من الطرق، كالتحديد بالتاريخ، أو التحديد باللغة، أو البحث في حدود النتائج، أو بالتلقيم المرتد للصلاحية أو الاتصال بالموضوع. وكانت نتائج عمليات البحث ترتب طبقيا في معظم الأحيان، بناء على خوارزمية ترتيب طبقي، عادة ما كان يتم الاحتفاظ بمكوناتها طي الكتمان. وقد تراجعت خيارات عدد المخرجات، وشكل العرض تدريجيا في هذا الجيل من نظم استرجاع الإنترنت، نظرا لتراجع أهميتها بالنسبة للمستفيد. وكنظم لاسترجاع المعلومات حقق الجيل الثاني قدرا كبيرا من التطور في الوظيفية والأداء.

ويمكن أن يكون الجيل الثالث لنظم استرجاع الإنترنت، قد أحرز المزيد من الارتقاء بمستوى الخواص الواردة في الجدول رقم ٨/ ٢؛ فلم تعد الوسائط المتعددة تعامل أولا بمعزل عن المعلومات النصية، في الجيل الجديد لنظم استرجاع الإنترنت، وإنها تكاملت تقنيات تجهيز الوسائط المتعددة واسترجاعها، مع تلك الخاصة بالمعلومات النصية. ونظرا لتميز العنكبوتية بقدرتها على تقديم المعلومات متعددة الوسائط، ولأنها قد أصبحت التطبيق الرئيس للإنترنت، فإنه لم يعد هناك مبرر لأن لا تكون هناك قدرة على تقديم الوسائط المتعددة، وتجهيزها واسترجاعها، جنبا إلى جنب مع المعلومات النصية، مسايرة لتطور البحث في استرجاع معلومات الوسائط المتعددة. ثانيا، تكاملت الأدلة ومحركات البحث، بدلا من توافرها ببساطة في الموقع نفسه، وبذلك يمكن للمستفيد أن يبحث في حدود فئة قابلة للتصفح (كما يتم بياهو!)، ثم تتجمع نتائج عمليات البحث في فئات للتصفح (كما في كلاستي Clusty). وفضلا عن تكشيف الكلمات المفتاحية، أصبح من الممكن لتكشيف المفاهيم أن يتم، بعد أن تحقق المزيد من التقدم في البحث حول معالجة اللغة الطبيعية والتعبير عنها. كذلك يمكن لمقومات البحث التي كانت تدعمها الأجيال السابقة، أن يرتفع مستواها وتتعزز، بالتركيز على الارتقاء بمستوى التحقيق. كما يتم أيضا إدخال مقومات بحث إضافية (كالبحث في اللغة الطبيعية على سبيل المثال) في رصيد مقومات الاسترجاع التي يوفرها النظام. وكذلك يمكن توجيه طريقة تعديل البحث بشكل أوضح، نحو تضييق عمليات البحث على نحو فعال؛ نظرا لأن المستفيد عادة ما تغمره الأعداد الهائلة للنتائج التي تسفر عنها نظم استرجاع الإنترنت. ويمكن لتقديم نتائج عمليات البحث أن يرتب طبقيا على نحو مثالي، وأن يطوع تبعا لمواصفات المستفيد، بدلا من خيارات النظام المحددة سلفا. فجوجل، على سبيل المثال، في واجهة بحثه المتطورة، يكفل للمستفيد أن يقرر من بين أشياء أخرى، ما إذا كان من الممكن للنتائج المزعجة أن يتم التخلص منها، أو ما إذا كان من الممكن

لصفحات في نطاق شبكة معينة أن تتاح. ويمكن لنظم استرجاع الإنترنت من الجيل الثالث، أن تكون قادرة على إبراز تطورات ملحوظة في أداء الاسترجاع.

وكما يتبين من هذه التوصيفات، فإن نظم استرجاع الإنترنت كما يبدو، قد تخطت جيلها الأول، وتتقدم بقوة في الجيل الثاني. وقد بدأ الجيل الثالث يلوح في الأفق، بإمكانيات قوية. ولقد لاحظنا منذ منتصف العقد الأخير من القرن العشرين، تغير أجيال نظم استرجاع الإنترنت، نظرا للتطور السريع للإنترنت، التي من أجلها صممت أدوات الاسترجاع تلك. ولا يمكن المبالغة في أهمية الإنترنت في مجال استرجاع المعلومات، وخصوصا في العصر الرقمي.

نظم استرجاع الإنترنت واسترجاع المعلومات:

لم تعد الإنترنت مستودعا عملاقا للمعلومات فحسب، وإنما أصبحت أيضا منصة مألوفة للوصول إلى أنواع أخرى من نظم استرجاع المعلومات. ونظم استرجاع الإنترنت أدوات لا غنى عنها ولاشك لاسترجاع المعلومات من الإنترنت. كذلك تعد هذه النظم بمثابة بوابات لمصادر المعلومات الأخرى. ونظرا لضخ المزيد والمزيد من المعلومات على الإنترنت، يمكن لأهمية نظم استرجاع الإنترنت أن تزداد تناسبيا.

وفضلا عن ذلك، فإن نظم استرجاع الإنترنت تتحول الآن إلى مختبر ونافذة عرض للأساليب الحديثة المتقدمة المتطورة لاسترجاع المعلومات، وهذا هو الدور الذي نهضت به نظم الخط المباشر في الماضي. ولكي يظل مواكبا لأحدث التطورات في مجال استرجاع المعلومات، يتعين على المرء مراقبة نظم استرجاع الإنترنت، بدلا من نظم الخط المباشر أو الأنواع الأخرى من النظم، عن كثب. فالتغيرات تحدث باستمرار، إن لم تكن على أساس يومي في نظم استرجاع الإنترنت. والنشرات الإخبارية القائمة على الإنترنت، وقد نشأت المواقع (مثل **Notess, 2008a; Third Door Media, 2008**) لإحاطة المستفيد علما بأحدث التطورات حول هذه النظم.

ومن ناحية أخرى، فإن نظم استرجاع الإنترنت كفئة، سيئة السمعة، نظرا لانخفاض معدل التحقيق في أدائها في البحث. وقد لا يكون من الممكن الوصول إلى المعلومات التي تسترجع من الإنترنت، نظرا لعوامل كالروابط المتقطعة والمتقادمة، وهذه نتيجة أقل احتمالا بكثير في البيئات الأخرى لاسترجاع المعلومات؛ كذلك يتعين على المستفيد أن يكون مسئولا أيضا عن الحكم على جودة المعلومات، نظرا لأن تقنية ضبط الجودة بالنسبة للمعلومات على الإنترنت، إن توافرت، فإنها في أحسن الأحوال مفككة البنيان إلى حد بعيد (كما هو الحال على سبيل المثال، حينما كانت فهرسة مصادر الإنترنت تتم للنظم القائمة على الأدلة).

والبحث، كما ذهب برور **Brewer (2001)**، هو أبرز وأهم جوانب الإنترنت، بعد الاتصال. ونظم استرجاع الإنترنت مصممة على وجه التحديد، لأغراض الاسترجاع. يضاف إلى ذلك أن هذه النظم يتم التعامل معها بلا رسوم بالنسبة للمستفيد النهائي إذا ما توافرت له فعلا فرصة التعامل مع الإنترنت. ومن ثم فإنه لا مبالغة في القول بأن نظم استرجاع الإنترنت تتبوأ مكانة بارزة في حقل استرجاع المعلومات، في الوقت الراهن، وسوف تظل كذلك لزمن طويل في المستقبل.

وعلى الرغم من أن نظم استرجاع الإنترنت تبدو واعدة، وبشكل استثنائي في حقل استرجاع المعلومات، فإن التحديات التي تواجهها يمكن أن تظل بحاجة إلى المزيد من الجهد **(Mayer, 2008)**. وتعتقد ماريسا ماير **Marissa Mayer**، نائب رئيس جوجل لمنتجات عمليات البحث وخبرات المستفيدين، أن البحث لا يتفق وقاعدة ٢٠-٨٠ المعتادة(*). وإنما بقدر ما يتعلق الأمر بجوجل، فإنه قد حل حوالي ٩٠

(*) ينص مبدأ باريتو **Pareto**، الذي يعرف أيضا بقاعدة ٢٠-٨٠، وقانون القلة الحيوية **vital few**، ومبدأ تشتت العوامل، على أنه بالنسبة لكثير من الوقائع، فإن حوالي ٨٠٪ من النتائج تتأتى من ٢٠٪ من الأسباب. وقد اقترح جوزيف إم. جوران **Joseph M. Juran**، استشاري=

بالمئة من مشكلات البحث، إلا أن العشرة بالمئة الأخيرة، من حلول مشكلات البحث تمثل أكثر من ٩٠ بالمئة من العمل. وتشمل أمثلة المشكلات في العشرة بالمئة الأخيرة، البحث المتنقل أو الجوال **mobile search** (بما في ذلك الأجهزة القابلة للارتداء أو الحمل **wearable**)، وبحث الأصوات، واللغة الطبيعية، وبحث الوسائط المتعددة الذي يتبع الطريقة القائمة على المحتوى، ونتائج البحث التي تقدم في وسائط (كالصور ومقاطع الفيديو) التي تعد أكثر ملاءمة للمستفيد. ومن هذا المنظور، لا يزال أمام نظم استرجاع الإنترنت كفئة، شوط طويل، يمكن أن تقطعه في حقل استرجاع المعلومات.

نظم استرجاع المعلومات: بعض الاتجاهات:

شهدت العقود الأخيرة نمو نظم استرجاع المعلومات، وتطورها في أبعاد كثيرة. وقد أحدث العصر الرقمي، وخصوصا تقنيات الشبكات، تغيرات في نظم استرجاع المعلومات، التي لم تعد تعمل بمعزل بعضها عن بعض، تبعا لنوعية النظم، وإنما لا تعمل هذه النظم بعضها عن بعض ، في حدود إحدى فئات نظم استرجاع المعلومات فحسب، (كنظم الخط المباشر مثلا) وإنما تندمج أيضا عبر خطوط حدودها. ومن ناحية أخرى تتيح تطبيقات الجيل الثاني للعنكبوتية الناشئ (كالتمييز الاجتماعي والموسوعات المصغرة **Wikis** على سبيل المثال) للمستفيدين القدرة على أن يصبحوا مساهمين في استرجاع المعلومات، ومن ثم فإنهم لا يكتفون بمجرد الانتظار على الطرف المتلقي في عملية تنظيم المعلومات واسترجاعها. ونكرس الأقسام الفرعية التالية لهذين الاتجاهين فيما يتعلق بنظم استرجاع المعلومات.

=إدارة الأعمال، هذا المبدأ، وأسماه باسم رجل الاقتصاد الإيطالي فلفريدو باريتو **Vilfredo Pareto**، الذي لاحظ عام ١٩٠٦، أن ٨٠٪ من الأراضي يمتلكها ٢٠٪ من السكان. (المترجم)

اندماج نظم استرجاع المعلومات:

يتبين من المعالجات والمناقشات في هذا الفصل، أن لكل نوع من نظم استرجاع المعلومات ملامحه ومقوماته الخاصة، ووظائفه وإمكانياته في استرجاع المعلومات، وإن كانت جميع هذه النظم مصممة، وتدار لأغراض استرجاع المعلومات. وقد بدأت مؤخرا مختلف نظم استرجاع المعلومات تندمج، من أجل توفير خدمات استرجاع معلومات أفضل وأكثر ملاءمة للمستفيدين. ويمكن للاندماج أن يتخذ أيا من الأشكال التالية.

والشكل الأول للاندماج، من الناحية التاريخية، هو تحميل نظم الأسطوانات المكتنزة على الفهارس المتاحة على الخط المباشر. ولما كانت هذه الأخيرة يتم إنشاؤها بوجه عام للمكتبات القائمة بذاتها، التي تقتني أيضا نظما للأسطوانات المكتنزة، يقوم كثير من المؤسسات بدمج هذين النظامين معا، بعد أن تؤدي الفئة الأخيرة رسالتها كاملة. ويتم توفير روابط من الفهرس المتاح على الخط المباشر إلى الأسطوانات المكتنزة، الأمر الذي يمكّن المستفيد من التعامل مع النظامين المختلفين لاسترجاع المعلومات في مكان واحد. كذلك تنشئ بعض المؤسسات روابط من فهارسها المتاحة على الخط المباشر، إلى نظم الخط المباشر، حتى يمكن التخلص من إجراءات الربط التي كان من الممكن أن تكون منفصلة عن بعضها البعض، عندما يكون المستفيد بصدد إجراء أي عمليات بحث على الخط المباشر. ويتم هذا الاندماج في الأساس عند مستوى الترابط بين النظم.

ويحدث الشكل الثاني للاندماج عند مستوى محتوى مراصد بيانات النظم، وذلك بين الفهارس المتاحة على الخط المباشر ونظم الخط المباشر. فمن الممكن على سبيل المثال، لعملية بحث في الفهرس المتاح على الخط المباشر، أن تبين أن وعاء بعينه لا وجود له في مجموعات المكتبة، وحينئذ يمكن للاستفسار عن ذلك الوعاء أن يقدم

لنظام الخط المباشر، الذي يرتبط به الفهرس المتاح على الخط المباشر، لمعرفة ما إذا كان الوعاء في نظام الخط المباشر أم لا. ومن المرجح إلى حد بعيد، أن تستخدم نظم الإمداد بالوثائق **document delivery** (مثل بروكوست **Proquest**) للحصول على نسخة من الوعاء. إلا أن الرابطة بين الفهارس المتاحة على الخط المباشر، ونظم الخط المباشر، هي التي تيسر فعلا عملية الاسترجاع.

ويحدث النوع الثالث من الاندماج عندما يتزايد استخدام العنكبوتية فائقة البناء أو التنظيم **hyperstructured** كمنصة لنظم استرجاع الإنترنت، إلى جانب استخدامها من أجل نظم الخط المباشر، ونظم الفهارس المتاحة على الخط المباشر. وتصبح نظم استرجاع المعلومات هذه مترابطة، بحيث يصبح ما بينها من حدود غير واضحة المعالم. فالحقل ٨٥٦ في مارك **MARC**- على سبيل المثال، وهو مواصفة معيارية للفهرسة- يتم تخصيصه لتسجيل الأماكن الموحدة للمصادر **URLs** الخاصة بمصادر الإنترنت. ويمكن للناتج النهائي لهذا الأسلوب أن يكون، كما بينا في القسم ٨/ ٤/ ٢/ ٢، فهارس متاحة على الخط المباشر، بالنسبة لمصادر الإنترنت. وفضلا عن ذلك هناك المزيد والمزيد من الكتب والدوريات التي تنشر إلكترونيا على العنكبوتية، وتتكامل مع مجموعات المكتبات، ويتم تمثيلها في الفهارس المتاحة على الخط المباشر. ومن ثم، فإن المعلومات المتوافرة على الإنترنت، التي كانت فيما مضى هدف نظم استرجاع الإنترنت بلا منازع، تشملها الفهارس المتاحة على الخط المباشر، التي كانت فيما مضى مخصصة كلية لتنظيم مجموعة المكتبات واسترجاعها. وقد أدى هذا الأسلوب إلى تحقيق الاندماج بين الفهارس المتاحة على الخط المباشر، ونظم استرجاع الإنترنت. ويمكن أن نجد مثالا آخر لمثل هذا الاندماج، حيثما تشتمل وثائق العنكبوتية على إشارات مرجعية أو استشهادات، تغطيها نظم الخط المباشر. وإذا افترضنا أن كانت إحدى وثائق العنكبوتية، إلى جانب استشهاداتها المرجعية، قد أمكن

العثور عليها بمساعدة أحد نظم استرجاع الإنترنت، فإنه يمكن للمستفيد أن يواصل البحث في نظام الخط المباشر، الذي يشتمل على الاستشهادات المرجعية المرتبطة بروابط فائقة. ويتحقق الترابط بين نظم الخط المباشر، ونظم استرجاع الإنترنت بسلاسة وبلا أدنى معوقات.

ومن منظور أكثر اتساعا، فإن العنكبوتية في سبيلها لأن تصبح المنصة بالنسبة للتعامل مع مختلف فئات نظم استرجاع المعلومات. ومن وجهة نظر أقل اتساعا، فإنه ليس من المستبعد بالنسبة لمستفيد ما، أن يتعامل مع منظومة معلومات مكتبة قائمة على العنكبوتية، تشمل جميع أنواع نظم استرجاع المعلومات؛ أي فهرس المكتبة المتاح على الخط المباشر، ونظم الخط المباشر التي تشترك فيها المكتبة، وبعض نظم استرجاع الإنترنت التي ترى المكتبة إتاحتها. فهل يمكن للعنكبوتية أن تصبح المنصة الوحيدة بالنسبة لنظم استرجاع المعلومات في المستقبل؟ هل يمكن لمنصة مشتركة أن تعمل على الحث على المزيد من الاندماج بين نظم استرجاع المعلومات؟ ما التأثير المحتمل للاندماج على المستفيد؟ لازلنا حتى الآن في انتظار إجابات لهذه الأسئلة.

الجيل الثاني للعنكبوتية ونظم استرجاع المعلومات:

لقد كان لظهور تطبيقات الجيل الثاني للعنكبوتية **Web 2.0**، كالموسوعات المصغرة **wikis**، والتمييز الاجتماعي، ولاشك، تأثيره على تنظيم المعلومات واسترجاعها بوجه عام، وعلى نظم استرجاع المعلومات على وجه الخصوص. ففي القسم ٢/ ١/ ٤ ورد الجيل الثاني للعنكبوتية بوصفه عنصرا في مناقشة التمييز الاجتماعي (أو التمييز بإيجاز) ودوره في التكشيف. وفي سياق نظم استرجاع المعلومات بدأت المدونات، والتمييز، والموسوعات المصغرة، تحدث تأثيرا ملحوظا على عملية استرجاع المعلومات.

فالمدونات عادة ما تستخدم في نطاق نظام استرجاع المعلومات، حتى يكون بإمكان المستفيد التعليق، على سبيل المثال، على الكتب التي اقتنتها إحدى المكتبات، أو المقالات المسترجعة من أحد نظم الخط المباشر. وتساعد رسائل المدونات من هذا النوع المستفيدين الأنداد في اتخاذ قرار ما إذا كان عليهم محاولة الاطلاع على مقالة بعينها، أو مراجعة كتاب معين. وقد تم وضع الموسوعات المصغرة في نظم استرجاع المعلومات (مثل Search wiki لجوجل) للسماح للمستفيدين بوضع لمساتهم الشخصية على ما يسترجعه ويقدمه نظام استرجاع المعلومات، وذلك بالارتقاء بالنتائج، وإضافة نتائج جديدة، وحذف نتائج قدمت من قبل، وتسجيل التعليقات **(Notess, 2008b)**. ويقودنا هذا التطبيق، على نحو ما، للاقتراب من تطويع طريقة تقديم نتائج عمليات البحث، كجزء من إضفاء الطابع الشخصي على البحث. وبناء على طبيعته الفريدة، يتزايد على نحو مطرد، إدخال التمييز في نظم استرجاع المعلومات، كوسيلة لتوفير نقاط وصول إضافية من وجهة نظر المستفيد. وقد بدأ ظهور نظم استرجاع المعلومات التي تنطوي على تطبيقات الجيل الثاني للعنكبوتية **Web 2.0**؛ فتو كولاب في إلزفير **2 collab at Elsevier**، على سبيل المثال، وهو شريك رئيس في نظم الخط المباشر، منصة لاسترجاع المعلومات، تكفل للمتعاملين معها الاستفادة من التمييز، والموسوعات المصغرة، وما يماثلها من تطبيقات العنكبوتية ٢.٠، أثناء التعامل مع نظمها على الخط المباشر (مثل ساينس دايركت **Science Direct**) وسوباكس **SOPACs (Social OPACs)**، وهي نظم فهارس متاحة على الخط المباشر بخواص العنكبوتية ٢.٠ **(Blyberg, 2008; Stephens, 2007)**، على الرغم من وجود قليل من التطبيقات.

وما يمكن أن يسفر عنه تكامل نظم استرجاع المعلومات وتقنيات العنكبوتية ٢.٠ فعلا، في حقل تنظيم المعلومات واسترجاعها، مجرد احتمالات. إلا أن هناك أمرا واحدا يتضح فعلا بجلاء؛ فالمستفيدون يبدون متحمسين فعلا، فيما يتعلق بقدرتهم على الإسهام في استرجاع المعلومات، بينما بدأ دارسو النظم واختصاصيو المعلومات النظر على نحو جاد في سبل جعل هذا التكامل ذا قيمة بالنسبة لكل الأطراف المشاركة.

المراجع

Antelman, Kristin, Lynema, Emily, and Pace, Andrew K. (2006). Toward a twenty-first century library catalog. *Information Technology and Libraries,* 25(3), 128-139.

Basch, Reva. (1993). Dialog's Rank command: Building and mining the data mountain. *ONLINE,* 17(4), 28-35.

Beaulieu, Micheline, and Borgman, Christine L. (Guest editor). (1996). Special topic: Current research in online public access systems. *Journal of the American Society for Information Science,* 47(7), 491-583.

Bharat, Krishna, and Hensinger, Monika R. (1998). Improved algorithms for topic distillation in a hyperlinked environment. In W. Bruce Croft, et al. (Eds.), *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 104-111). New York: Association for Computing Machinery.

Blyberg, John. (August 16, 2008). SOPAC 2.0: What to expect. Retrieved October 7, 2009, from www.blyberg.net/2008/08/16/sopac-20-what-to-expect.

Borgman, Christine L. (1986). Why are online catalogs hard to use? Lessons learned from information retrieval studies. *Journal of the American Society for Information Science,* 37, 387-397.

Borgman, Christine L. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science,* 47(7), 493-503.

Bourne, Charles P. (1980). On-line systems: History, technology, and economics. *Journal of the American Society for Information Science,* 31(3), 155-160.

Bourne, Charles P., and Hahn, Trudi Bellardo. (2003). *History of online information services 1963-1976.* Cambridge, MA: MIT Press.

Brewer, Eric A. (2001). When everything is searchable. *Communications of the ACM,* 44(3), 53-55.

Brin, Sergey, and Page, Lawrence. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems,* 30 (1-7), 107-117.

Chowdhury, Gobinda G. (1999). *Introduction to modern information retrieval.* London: Library Association Publishing.

Chu, Heting. (1998). Internet search services vs. online database services. In Martha E. Williams (Ed.), *Proceedings of the 19th National Online Meeting* (pp. 69-75). Medford, NJ: Information Today.

Chu, Heting. (2005). Taxonomy of inlinked web entities: What does it imply for webometric research? *Library & Information Science Research,* 27(1), 8-27.

Fallows, Deborah. (2008). Search engine use. Pew Internet & American Life Project. Retrieved October 7, 2009, from www.pewinternet.org/-/media/Files/Reports/2008/P2P_Search_Aug08.pdf.pdf

Fallows, Deborah, Rainie, Lee, and Mudd, Graham. (2004). The popularity and importance of search engines. Pew Internet & American Life Project. Retrieved October 7, 2009, from www.pewinternet.org/-/media/Files/Reports/2004/P2P_Data_Memo_Searchengines.pdf.pdf

Festa, Paul. (February 26, 2001). Search project prepares to challenge Google. Retrieved October 7, 2009, from news.cnet.com/2/00-1023-253141.html

Frauenfelder, Mark (September 25, 1998). The future of search engines. *The Industry Standard: The Newsmagazine of the Internet Economy.* Retrieved May 26, 1999, from www.thestandard.com/articles/ article_print/0,1454,1826,00.html

Google. (2008a). About Google Book Search. Retrieved December 8, 2008, from books.google.com/intl/en/googlebooks/history.html

Google. (2008b). Technology overview. Retrieved November 30, 2008, from www.google.com/corporate/tech.html

Hahn, Trudi Bellardo. (1996). Pioneers of the online age. *Information Processing & Management,* 32(1), 33-48.

Harter, Stephen P. (1986). *Online information retrieval: Concepts, principles, and techniques.* New York: Academic Press.

Hildreth, Charles R. (1984). Pursuing the ideal: Generations of online catalogs. In B. Aveney and B. Butler (Eds.), *Online catalogs, online reference: Converging trends. Proceedings of a Library and Information Technology Association Preconference Institute* (pp. 31-56). Chicago: American Library Association.

Hildreth, Charles R. (1985). Online public access catalogs. *Annual Review of Information Science and Technology,* 20, 233-285.

Hildreth, Charles R. (1988). Online library catalogues as information retrieval systems: What can we learn from research? In P. A. Yates-Mercer (Ed.), *Future trends in information science and technology. Proceedings of the Silver Jubilee Conference of the City University's Department of Information Science* (pp. 9-25). London: Taylor Graham.

Hildreth, Charles R. (1995). Online catalog design models: Are we moving in the right direction? A report submitted to the Council on Library Resources. (Updated March 27, 2000). Retrieved December 6, 2008, from myweb.cwpost.liu.edu/childret/clr-opac.html

Kleinberg, Jon. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM,* 46(5), 604-632.

Large, Andrew, Tedd, Lucy A., and Hartley, R. J. (1999). *Information seeking in the online age: Principles and practice.* London: Bowker-Saur.

Lawrence, Steve, and Giles, C. Lee. (April 3, 1998). Searching the World Wide Web. *Science,* 280(5360), 98-100.

Lawrence, Steve, and Giles, C. Lee. (July 8, 1999). Accessibility of information on the web. *Nature,* 400, 107-109.

LeVan, Ralph. (2003). Z39.50 as a web service. [Presentation slides]. Retrieved December 1, 2008, from staff.oclc.org/-levan/docs/srw-niso20030430.ppt

Lewandowski, Dirk. (2005). Web searching, search engines and information retrieval. *Information Services & User,* 25, 137-147.

Malinconico, S. Michael. (1984). Catalogs & cataloging: Innocent pleasures and enduring controversies. *Library Journal,* 109(11), 1210-1213.

Marcum, Deanna B. (2006). The future of cataloging. *Library Resources & Technical Services,* 50(1), 5-9.

Markey, Karen. (2007). The online library catalog: Paradise lost and paradise gained? *D-Lib Magazine*, 13(1/2). Retrieved December 6, 2008, from dlib.org/dlib/january07/Markey/01/Markey.html

Mayer, Marissa. (September 10, 2008). The future of search. Retrieved December 9, 2008, from googleblog.blogspot.com/2008/09/future-of-search.html

McKiernan, Gerry. (2001). Beyond bookmarks: Schemes for organizing the web. Retrieved January 19, 2009, from www.public.iastate.edu/-CYBERSTACKS/CTW.htm

Meadow, Charles T., Boyce, Bert R., and Kraft, Donald H. (1999). *Text information retrieval systems.* Orlando, FL: Academic Press.

Nahl-Jakobovits, Diane, and Tenopir, Carol. (1992). Databases online and on CD-ROMs: How do they differ? Let us count the ways. *Database,* 15(1), 42-50.

Notess, Greg R. (2008a). Search engine showdown: The users' guide to web searching. Retrieved December 11, 2008, from www.searchingineshowdown.com

Notess, Greg R. (December 4, 2008b). Customize your own Google results with Search Wiki. ITI NewsLink. Retrieved December 9, 2008, from newsbreaks.infotoday.com/nbReader.asp?ArticleId=51802

Rowley, Jennifer, and Slack, Frances. (1997). The evaluation of interface design on CD-ROMs. *Online and CDROM Review,* 21(1), 3-11.

Schwartz, Candy. (2000). Meeting review: Notes from the Boston 2000 Search Engine Meeting. *Bulletin of the American Society for Information Science,* 26(6), 26-28.

Smith, Ian. (2000). The invisible web: Where search engines fear to go. Retrieved December 12, 2000, from www.powerhomebiz.com/ vol25/invisible.htm

Stephens, Michael. (2007). Web 2.0 & libraries, part 2: Trends and technologies. Chapter 3, Technology trends for a 2.0 world. *Library Technology Reports,* 43(5), 32-44.

Tech Target. (2005). Digital versatile disk. Retrieved December 6, 2008, from whatis.techtarget.com

Tech Target. (2007). Vortal. Retrieved December 9, 2008, from whatis.techtarget.com

Third Door Media. (2008). Search engine land. Retrieved December 11, 2008, from searchengineland.com

Walker, Geraldene, and Janes, Joseph. (1999). *Online retrieval: A dialogue of theory and practice* (2nd ed.). Englewood, CO: Libraries Unlimited.

الفصل التاسع

استرجاع المعلومات الفريدة في محتواها أو في شكلها

لقد استكشفنا حتى الآن موضوع استرجاع المعلومات، مفترضين أن المعلومات في المقام الأول نصوص بالإنجليزية، ولها بنيان خطي. ويركز هذا الفصل على استرجاع المعلومات الفريدة في محتواها أو في شكلها، أي المعلومات متعددة اللغات، ومعلومات الوسائط المتعددة، ومعلومات النصوص الفائقة أو الوسائط الفائقة.

المعلومات متعددة اللغات

هناك ما يقارب من ٦٩١٢ لغة حية في العالم، من بينها ٢٠٩٢ تستعمل في أفريقيا، و١٠٠٢ في الأمريكتين، و٢٢٦٩ في آسيا، و٢٣٩ في أوروبا، و١٣١٠ في المحيط الهادئ (Gordon, 2005, p. 15). وبناء على تقدير أعد في فبراير عام ١٩٩٩، كانت اللغات الأولى الثلاث المتربعة على القمة، في العالم في ذلك الوقت، هي الصينية التقليدية المبسطة **Mandarin Chinese** (٨٨٥ مليون ناطق بها)، والإسبانية (٣٣٢ مليون ناطق بها)، والإنجليزية (٣٢٢ مليون ناطق بها) **(Sands, 2000)**. وبينما يمكن لبشر في حياتهم اليومية، الاعتماد على لغتهم الأم دون سواها، فإن العالم أحادي اللغة زائل حتما في هذا العصر الرقمي، بقدر ما يحظى تنظيم المعلومات واسترجاعها بالاهتمام؛ نظرا لأن المعلومات يتم تسجيلها وتبادلها بلغات متعددة، في شتى أنحاء

العالم، بما في ذلك الإنترنت. ومن ثم فإن الحاجة إلى تنظيم واسترجاع المعلومات متعددة اللغات، تصبح ضرورية، حتى وإن كان من الممكن للغة واحدة أن تكون لها السيطرة في وسط معين، إذا ما كانت تلك اللغة هي ما يستعملها معظم البشر هناك. وبعبارة أخرى فإن بيئة تنظيم المعلومات واسترجاعها برمتها، ينبغي أن تكون متعددة اللغات، لتلبية الاحتياجات المعلوماتية لمختلف فئات المستفيدين الذين يتحدثون لغات مختلفة.

ولقد كانت الإنجليزية هي اللغة الرئيسة للاتصال العلمي، منذ الحرب العالمية الثانية، ولا تزال في الصدارة بين جميع اللغات المستعملة على الإنترنت حتى الآن. ومع ذلك فإن الفجوة بين الإنجليزية واللغات الأخرى (كالصينية والإسبانية على سبيل المثال)، تتراجع بسرعة في أوساط المتعاملين مع الإنترنت. وكما ذهبت مجموعة مينيواتس للتسويق **Miniwatts Marketing Eroup (2009)**، على سبيل المثال، فإن لغات القمة الثلاث التي تستعمل على الإنترنت، كانت الإنجليزية (٢٨.٧ بالمئة)، والصينية (٢١.٧ بالمئة)، والإسبانية (٨.٠ بالمئة)، وذلك في يونيو عام ٢٠٠٩. وفي مقابل ذلك كانت النسبة المئوية للمعلومات بالإنجليزية أعلى بكثير، عندما انطلقت الإنترنت جماهيريا في مطلع العقد الأخير من القرن العشرين. ويدل هذا التغير على حقيقة أن الإنترنت في سبيلها على الأقل، لأن تصبح متعددة اللغات على نحو أكثر.

وينبغي للمعلومات متعددة اللغات أن يتم ترميزها للتجهيز بواسطة الحاسبات، قبل أن يكون من الممكن تنظيمها واسترجاعها رقميا، واليونيكود **Unicode** مواصفة معيارية وضعت خصيصا لمثل هذه الأغراض، بحيث يكون من الممكن ترميز أو تكويد كل حرف أو رمز في أي لغة، برقم فريد بالبايت، بصرف النظر عن المنصة أو التطبيق (**Unicode, Inc.,2008**). وفي زمن ما قبل اليونيكود، كان لكل لغة نظام التكويد الخاص بها، على نحو يجعل التشغيل التبادلي فيما بينها ضربا من

المستحيل. فقد كان الآسكي ASCII (الكود المعياري الأمريكي لتبادل المعلومات American Standard Code for Information Interchange) الشهير، مصمما لتكويد الهجائية الإنجليزية، ولكنه لا يستطيع معالجة الأحرف في اللغات الأخرى (كالصينية مثلا). كما اتضح أيضا أن اليونيكود ضرورة بالنسبة لتطور الإنترنت؛ نظرا للمعلومات متعددة اللغات، التي تعالجها الشبكة الكونية معالجة الأحرف في اللغات الأخرى (كالصينية مثلا). كما اتضح أيضا أن اليونيكود ضرورة بالنسبة لتطور الإنترنت؛ نظرا للمعلومات متعددة اللغات، التي تعالجها الشبكة الكونية (Lerner, 2003). واليونيكود ليس ضروريا في تكويد المعلومات متعددة اللغات للمزيد من التجهيز الرقمي فحسب، وإنما يكفل لمتصفح العنكبوتية القدرة على اكتشاف ما إذا كانت صفحة العنكبوتية الواردة أمامه، تستخدم إحدى خطط التكويد المركبة فعلا على حاسب المتصفح (Coyle, 2005). ودعما لهذه الوظيفة الأخيرة يلغي اليونيكود حاجة المتعامل لأن ينتقل يدويا من إحدى خطط التكويد (كالإنجليزية مثلا) إلى أخرى (كالصينية مثلا) عندما يكون بصدد النظر في معلومات متعددة اللغات كنتائج إحدى عمليات البحث مثلا.

استرجاع المعلومات متعددة اللغات في الماضي:

قبل ظهور الإنترنت، ونظم استرجاع الإنترنت، كانت المعلومات متعددة اللغات تدخل في نظم الخط المباشر، ونظم الأسطوانات المكتنزة، ونظم الفهارس المتاحة على الخط المباشر، على أساس انتقائي. وعادة ما تتبع هذه النظم سياسة للانتقاء فيما يتعلق بتجميع المعلومات متعددة اللغات؛ فقد كان من الممكن على سبيل المثال، توفير مقومات الترجمة أو النقحرة Transliteration، إذا كان من المزمع إدخال لغة معينة. وفي نظم استرجاع المعلومات متعددة اللغات، لم يكن العثور على المعلومات باستعمال اللغة نفسها التي سجلت بها المعلومات (كالبحث عن المعلومات الصينية

باستعمال استفسارات بالصينية على سبيل المثال) فيما يبدو يحظى بالدعم على نطاق واسع، على الرغم من أنه كانت هناك بعض الاستثناءات. ويمكن القول بوجه عام أن المعلومات باللغات الأجنبية كانت تنقحر لأغراض الاسترجاع في الفهارس المتاحة على الخط المباشر. وغالبا ما تستعمل نظم الخط المباشر، ونظم الأسطوانات المكتنزة، لغة أولى لتنظيم معلومات اللغات الأجنبية (في التكشيف والاستخلاص على سبيل المثال)، وكان من الممكن استرجاعها فيما بعد باللغة الأولى نفسها.

وهناك نظم مصممة لاسترجاع المعلومات متعددة اللغات مباشرة، بدلا من استرجاعها عن طريق الترجمة والنقحرة؛ فالفهارس المتاحة على الخط المباشر، في الصين على سبيل المثال، عادة ما تدعم طريقتين لغويتين، إحداهما للغة الصينية، والأخرى لما يسمى اللغات الغربية، كالإنجليزية والفرنسية مثلا. وبإمكان المتعاملين استعمال الصينية على سبيل المثال، لاسترجاع المعلومات بالصينية، والإنجليزية لاسترجاع المعلومات بالإنجليزية. وكان من المفترض في نظم استرجاع المعلومات متعددة اللغات، أن تسفر عن نتائج أفضل من تلك التي تعتمد على الترجمة والنقحرة، لأن النظم متعددة اللغات بإمكانها تجنب إدخال المزيد من مظاهر التفاوت أو الاختلاف، التي يمكن أن تنجم عن الترجمة أو النقحرة، في عملية الاسترجاع.

استرجاع المعلومات متعددة اللغات على الإنترنت:

على الرغم من أن نظم الخط المباشر، والأسطوانات المكتنزة، والفهارس المتاحة على الخط المباشر، تجمع معلومات متعددة اللغات، فإن الحاجة إلى استرجاع المعلومات متعددة اللغات في تلك البيئات، لا تبدو قوية كما هي على الإنترنت، نظرا للعوامل التالية: أولا، أن تغطية المعلومات متعددة اللغات في تلك النظم، كانت انتقائية إلى حد بعيد، كما تقتصر على المعلومات عالية الجودة. ثانيا، أن الإنترنت شبكة كونية للشبكات، ومن الطبيعي أن تحمل معلومات متعددة اللغات، أكثر بكثير من الأنواع الأخرى من نظم

استرجاع المعلومات. ثالثا، يسهم غياب تدابير ضبط الجودة، كما بينا في القسم ٨/ ٤/ ٢/ ١ كذلك في ازدهار المعلومات متعددة اللغات على الإنترنت.

ومما لاشك فيه، أن الإنترنت، كبيئة لاسترجاع المعلومات، تختلف عن البيئات الأخرى لاسترجاع المعلومات، نظرا لتعدد فئات أو اختلاف خصائص المتعاملين معها. وقد تكهن فالوز **Fallows (2007)**، عند الكتابة لمشروع إنترنت المقصورة والحياة الأمريكية **Pew Internet and American Life Project**، أن يكون من الممكن للصين أن تحل محل الولايات المتحدة الأمريكية، في إجمالي عدد المتعاملين، في غضون سنوات قليلة. وقد تأكد هذا التكهن أسرع مما كان متوقعا، من جانب مركز معلومات شبكة الإنترنت الصينية **China Internet Network Information Center (CNNIC)**، عندما أعلن أن المتعاملين مع الإنترنت في الصين بلغ إجمالي عددهم ٢٩٨ مليون، في نهاية ديسمبر عام ٢٠٠٨ **(CNNIC, 2008)**. ويدل تقرير مركز معلومات شبكة الإنترنت الصينية، إلى جانب البيانات الواردة من مجموعة مينيواتس للتسويق **(Miniwatts Marketing Group, 2009)**، فيما يتعلق بعدد المتعاملين مع الإنترنت في دول كاليابان، وفرنسا، وألمانيا، يدل على أن الإنترنت تغادر كما يبدو ماضيها أحادي اللغة. وتتعامل نظم استرجاع الإنترنت مع قضية استرجاع المعلومات متعددة اللغات، إما بدعم عمليات البحث متعدد اللغات، وإما بتوفير مقومات الترجمة التي طورتها شركات مثل سيستران **(www.systransoft.com) SYSTRAN**. كما كان لآلتافيستا **AltaVista** الريادة في كل من المجالين، بالتعاون مع سيستران، بينما حذت كثير من نظم استرجاع معلومات الإنترنت الأخرى (مثل جوجل) حذوه، وخصوصا في مجال البحث متعدد اللغات.

وعندما بدأ تطور نظم استرجاع الإنترنت، كانت الإنجليزية هي اللغة الوحيدة التي يمكن للمتعاملين استعمالها للاسترجاع. وبمقدور هؤلاء المتعاملين البحث الآن

في الإنترنت، باستعمال أكثر من اثنتي عشرة لغة مختلفة، تشمل الإنجليزية، والصينية، والفرنسية، والألمانية، والإسبانية، والروسية. ومن الممكن إجراء البحث متعدد اللغات في الإنترنت بطريقتين؛ الطريقة الأولى هي إجراء البحث باستعمال أحد نظم استرجاع الإنترنت المصمم خصيصا للغة معينة (مثل ياهو! في فرنسا) وذلك باستفسارات بحث بتلك اللغة. والطريقة الثانية هي قصر نتائج البحث على لغة بعينها، بانتقاء اللغة التي يقع عليها الاختيار، الواردة على الصفحة الرئيسة لأحد نظم استرجاع الإنترنت. ويظل استفسار البحث بالإنجليزية على سبيل المثال، إلا أن النتائج يمكن أن تكون باللغة المحددة (كالإسبانية مثلا). وينبغي للاسترجاع من هذا النوع، أن ينظر إليه، بالمعنى الدقيق، بوصفه استرجاعاً للمعلومات عبر اللغات، الذي نعالجه في القسم ٩/ ١/ ٣.

وفضلا عن البحث متعدد اللغات، يمكن الحصول على المعلومات متعددة اللغات على الإنترنت بواسطة الترجمة، وذلك بفصل الجهود المشتركة التي تبذلها نظم استرجاع الإنترنت، وشركات برمجيات الترجمة الآلية. وتضافر الجهود بين آلتافيستا **Altavista** وسيستران **SYSTRAN** مثال على ذلك؛ فمن الممكن على سبيل المثال، ترجمة أي نتائج يتم استرجاعها بواسطة آلتافيستا، من الإنجليزية، إلى الصينية، أو الفرنسية، أو الألمانية، أو الإيطالية، أو اليابانية، أو الكورية، أو الإسبانية، أو البرتغالية، والعكس صحيح، وذلك باستعمال برمجيات سيستران، وإن كانت جودة الترجمة ليست مرضية تماما؛ نظرا لأوجه القصور الكامنة في الترجمة الآلية **(Balkin, 1999)**.

البحث حول استرجاع المعلومات متعددة اللغات:

اجتذب استرجاع المعلومات متعددة اللغات اهتمام الباحثين في السنوات الأخيرة، وأصبح مساراً (أي مشكلة فرعية بعينها للاسترجاع في ترِك ٤ حتى ترِك ٦ **(TREC-4 through TREC-6**). وقد تناول المسار متعدد اللغات بالدرس، قضايا

استرجاع المعلومات بلغات غير الإنجليزية. وقد تمت دراسة الإسبانية والصينية، اللغتين غير الإنجليزيتين المستخدمتين على أوسع نطاق (**Gordon, 2005**)، من جانب أكثر من عشرين فريقا للبحث، من عام ١٩٩٥ حتى عام ١٩٩٧. وتبين للباحثين (**Spärck Jones, 2000**) أن:

> يبدو الامتداد إلى الإسبانية متعددة اللغات متواضع نسبيا، أما إلى الصينية فأكثر إلحاحا لتغطية نظام كتابة يختلف اختلافا جوهريا. ومن ثم فإنه يثير قضية ما إذا كان في التطبيق، في مقابل المبادئ الأساس، أن طرق التكشيف والبحث التي تطورت للإنجليزية، كتلك التي تستخدم الأساليب الإحصائية، يمكن أن تنتقل إلى لغات أخرى مختلفة تماما، وأن تكون فعالة بالنسبة لتلك اللغات. (**P. 46**)

وتتطلب اللغات المختلفة بالفعل، نظم استرجاع معلومات بإمكانيات مختلفة، لمعالجة خصائصها المميزة. فمن المهام المثيرة للتحديات إلى حد بعيد، على سبيل المثال، تحليل النص الصيني آليا إلى مصطلحات استفسار لها دلالتها؛ نظرا لأنه لا يوجد فراغ بين الكلمات في الكتابة الصينية إلا في نهاية الجملة. ومن ثم، قارنت معظم فرق البحث المشاركة في مسار الصينية متعددة اللغات في ترِك -٦ **TREC-6**، طرق تحليل مختلفة (**Voorhess & Harman, 2000**). والبحث بمراعاة شكل الحرف الأول، وهو من الأساليب الرئيسة لاسترجاع المعلومات في الإنجليزية، غير قابل للتطبيق في الصينية، نظرا لأنه ليست هناك حروف كبيرة أو صغيرة في النص الصيني.

والمجال المتصل بالاسترجاع متعدد اللغات، هو البحث عبر اللغات **Cross-language**، الذي تتوزع فيه الوثائق على عدة لغات (كالإنجليزية، أو الألمانية، أو الفرنسية، أو الإيطالية على سبيل المثال) وتصاغ الاستفسارات بكل لغة من هذه

اللغات. وقد أضيف مسار استرجاع المعلومات عبر اللغات **Cross-Language Information Retrieval (CLIR)** إلى سلسلة مؤتمرات استرجاع النصوص **Text REtrieval Conference (TREC)**، في عام ١٩٩٧، بدءا بترِك-٦ **TREC-6** حتى ترِك ٢٠٠٢ **TREC 2002**. والهدف الرئيس لمسار استرجاع المعلومات عبر اللغات، هو استرجاع الوثائق بإحدى اللغات المستهدفة، باستعمال استفسارات مكتوبة بلغة أخرى مختلفة. وفي ترِك-٦، جربت ثلاث عشرة مجموعة مشاركة في مسار استرجاع المعلومات عبر اللغات، ثلاثة طرق رئيسة للاسترجاع عبر اللغات، هي: (١) الترجمة الآلية، التي كانت تترجم بها الاستفسارات أو الوثائق إلى اللغة المستهدفة، (٢) استعمال معاجم ثنائية اللغة قابلة للقراءة بواسطة الحاسبات، أو أي مصادر لغوية أخرى قائمة، و(٣) استعمال رصيد من المصادر لاختبار طريقة الاسترجاع عبر اللغات أو لتمكينها**(Voorhess & Harman, 2000)**. وفي حدود ما يهم القضايا التقنية لاسترجاع المعلومات عبر اللغات **CLIR**، استعرض كيشيدا **Kishida (2005)** تفصيلا، إجراءات الترجمة، وطرق فك غموض الترجمة، واستراتيجيات المضاهاة، ضمن جهود أخرى كثيرة بذلها الباحثون في استرجاع المعلومات عبر اللغات على مر السنين.

وقبل أن ينتهي ترِك **TREC** من مسار استرجاع المعلومات عبر اللغات **CLIR**، في عام ٢٠٠٢، ظهرت منصتان رئيستان أخريان لاسترجاع المعلومات عبر اللغات؛ والمنصة الأولى هي منتدى التقييم عبر اللغات **Cross-Language Evaluation Forum (CLIF;www.clef-campaign.org)** الذي يهدف إلى تشجيع البحث والتطوير في الوصول إلى المعلومات متعددة اللغات، من المنظور الأوروبي في المقام الأول، والمنصة الثانية هي مجموعة اختبار المعهد الوطني للمعلوماتية لورش عمل نظم استرجاع المعلومات **National Institute of Informatics, Test Collection for IR Systems (NTCIR) Workshop**،

ومقرها باليابان (**research.nii.ac.jp/ntcir/online/prop-en.html**). ومحاكاة لسلسلة ترِك **TREC**، صممت ورش عمل نظم استرجاع المعلومات **NTCIR Workshops** لتشجيع البحث في تقنيات الوصول إلى المعلومات، بما في ذلك استرجاع المعلومات. واسترجاع المعلومات عبر اللغات، أحد المهام البحثية التي تنهض بها ورش عمل مجموعة اختبار نظم استرجاع المعلومات بالمعهد الوطني للمعلوماتية **NTCIR**. وتشمل اللغات غير الإنجليزية التي تحظى بالاهتمام بانتظام في ورش العمل تلك، اليابانية، والصينية، والكورية. ويعمل كل من منتدى التقييم عبر اللغات **CLEF**، ومجموعة اختبار المعهد الوطني للمعلوماتية **NTCIR** بهمة ونشاط، على استكشاف المشكلات التي يعاني منها استرجاع المعلومات عبر اللغات. وعلى الرغم من تأسيس كل من مؤتمرات استرجاع النصوص **TREC**، ومنتدى التقييم عبر اللغات **CLEF**، ومجموعة اختبار المعهد الوطني للمعلوماتية **NTCIR**، يتم الإعراب عن المخاوف فيما يتعلق بالاهتمام الملائم بكثير من اللغات التي لا تحظى بالتغطية في تلك المشروعات، وتوسعة العنكبوتية متعددة اللغات، بالإضافة إلى الإنتاج الفكري خلاف المعلومات الإخبارية (**Gey, Kando & Peters, 2005**). وللتعامل مع تلك المخاوف، يحاول باحثو مسار استرجاع المعلومات عبر اللغات **CLIR**، على نحو متزايد، اختبار مثل هذه الطرق الخاصة بمعالجة اللغة الطبيعية، كالتقسيم الجماعي **Copound splitting**، والترجمة الآلية (من أمثال **Chen & Gey, 2004, Hollink, et al., 2004; Si, et al., 2008**). ويبدو أن الذكاء الاصطناعي ينهض بدور في هذه الجهود.

وقصارى القول، فإن البحث متعدد اللغات، يعني أن تكون المعلومات المستهدفة، وكذلك الاستفسارات باللغة نفسها، بينما يعني الاسترجاع عبر اللغات أن تكون المعلومات المستهدفة، والاستفسارات بلغات مختلفة. وقصر نتائج البحث في الإنترنت على لغة واحدة بعينها، كما ذكرنا في ٩/ ١/ ٢، في واقع الأمر، استرجاع عبر

اللغات؛ نظرا لأن الاستفسار ينبغي أن يترجم من إحدى اللغات إلى لغة المعلومات المستهدفة، قبل إجراء عملية البحث. وسواء كان الاسترجاع متعدد اللغات، أو عبر اللغات، فإن المتعاملين يتعين عليهم دائما الانتباه إلى خصائص اللغة التي سجلت بها المعلومات؛ نظراً لاختلاف كل لغة عن غيرها. ولا يمكن إلا من خلال هذا المنظور إنجاز الاسترجاع متعدد اللغات أو عبر اللغات بنجاح.

المعلومات متعددة الوسائط:

بناء على دراسة أجراها معهد إدارة ونظم المعلومات، بجامعة كاليفورنيا في بركلي **School of Information Management and Systems, University of California at Berkely (Lyman, et al., 2000)**، ينتج العالم ما بين ١-٢ إكزابايت **exabytes** (الإكزابايت الواحدة = بليون جيجا بايت) من المعلومات المختلفة سنويا. وعلى الرغم من أن تقرير الدراسة لم يحدد النسبة المئوية للمعلومات متعددة الوسائط من المجموع، فإن مقدار المعلومات متعددة الوسائط (كالنشرات الإخبارية، ومؤتمرات الفيديو، وبيانات الاستكشاف، والتسجيلات السمعبصرية للقاءات المؤسساتية، والمحاضرات الدراسية، على سبيل المثال) ينمو بلا شك بشكل كبير؛ نظرا لأن العنكبوتية القابلة للتكشيف على الصعيد العام، وحدها كانت تشتمل على ما يقدر بمئة وثمانين مليون صورة، أي ثلاث تيرابايت(*) **terabytes** من بيانات الصور، في فبراير عام ١٩٩٩ **(Lawrence & Giles, 1999)**. ولما كانت العنكبوتية تواصل نموها السريع، وتحمل المزيد من المعلومات متعددة الوسائط، فإن الحاجة إلى العثور على هذه المعلومات والحصول عليها، تزداد أيضا على نحو مطرد.

(*) التيرابايت تساوي تريليون بايت، أي ألف جيجا بايت. (المترجم)

وقد عُرِّفت الوسائط المتعددة، في الفصل الثالث، بوصفها أي جمع بين المعلومات الصوتية والمصورة والنصية، ويشتمل الشكل رقم ٣/ ١ على التقسيم الشجري للمعلومات متعددة الوسائط. وينبغي أن نلاحظ أن المصطلح "صورة" **image** يستعمل أحيانا على نحو فضفاض، للدلالة على جميع أنواع الوسائط المتعددة؛ نظرا لأن كلا من الصور الثابتة والصور المتحركة، يمكن أن تشتمل على معلومات نصية، بينما يمكن للصور المتحركة أن تشتمل أيضا على معلومات صوتية. ولما كانت المعلومات النصية هي موضوع المناقشة في معظم أجزاء هذا الكتاب، فإن التركيز في هذا القسم ينصب على المعلومات غير النصية، التي تتكون من الصور الثابتة، والأصوات، والصور المتحركة.

وقد تطورت طرق وخوارزميات عدة لاسترجاع المعلومات متعددة الوسائط، إلا أنها تنقسم إلى فئتين عريضتين؛ هما النهج القائم على الوصف **description-based approach**، والنهج القائم على المحتوى **content-based approach**. ويستخدم النهج القائم على الوصف، ببساطة، يدويا، العناوين الفرعية، والكلمات المفتاحية، أو غيرها من الواصفات (كاسم الفنان، وحجم العمل على سبيل المثال) الخاصة بالمعلومات متعددة الوسائط، لأغراض الاسترجاع. ويتدخل البشر بشكل مباشر في عملية الاسترجاع، عند اتباع هذا النهج. ومن ناحية أخرى يدل النهج القائم على المحتوى، على الأساليب التي تستند إلى التجهيز الآلي للمعلومات متعددة الوسائط نفسها.

ويتم اتباع النهج القائم على الوصف، والنهج القائم على المحتوى، من جانب فئتين مختلفتين من الباحثين **(Cawkell, 1992; Chu, 2001; Persson, 2000)**. ويبدو أن من ينتمون إلى مجال علم الحاسبات يركزون على النهج القائم على المحتوى، بينما تركز أوساط علم المعلومات، بما في ذلك علم المكتبات، على النهج القائم على

الوصف (Cawkell, 1992; Rasmussen, 1997). ويأتي هذا التقسيم نتيجة لعاملين؛ أولهما اختلاف الخلفية الأكاديمية، وكذلك مقومات التدريب لكل من الفئتين من الباحثين؛ فأولاهما في علم الحاسبات والثانية في علم المعلومات. أما العامل الثاني فهو اختلاف طرق التجهيز التي تتبع بالنسبة لاسترجاع المعلومات متعددة الوسائط، في كل من النهجين؛ إذ أن النهج الأول آلي، والثاني يدوي. وتعالج الأقسام التالية كيف يتم تطبيق كل من النهجين في استرجاع الأنواع الثلاثة للمعلومات متعددة الوسائط، وهي الصور الثابتة، والأصوات، والصور المتحركة.

استرجاع الصور الثابتة:

تشمل الصور، كما بينا في القسم ٣/ ٣/ ١ كلا من الصور الثابتة والصور المتحركة. وتشمل الصور الثابتة الرسوم والصور الضوئية، والملصقات **posters**، وصور الوثائق (مثل ملفات صيغ الوثائق متعددة المنصات **Portable Document Format (PDF)**) وما شابه ذلك. ومن الممكن تقسيم الصور المتحركة، التي تشمل الصور الثابتة بالإضافة إلى الصور المتحركة، إلى فئتين فرعيتين؛ الصور المتحركة المصحوبة بالصوت، التي تشكل أشرطة الفيديو أو الأفلام، والصور المتحركة بلا صوت، وهي الرسوم المتحركة أو الأفلام الصامتة. ولأغراض التوضيح، نناقش في هذا القسم الصور الثابتة فقط، بينما نناقش الصور المتحركة في القسم ٩/ ٢/ ٣.

للصور الثابتة معناها الخاص وخصائصها التي تشكل الأسس بالنسبة لمهام الأداء أو الاسترجاع. وفيما يتعلق بفنون عصر النهضة، يحدد بانوفسكي **Panofsky (1955)** معالم ثلاث طبقات للمعاني بالنسبة للصور. والطبقة الأولى هي ما قبل صناعة التماثيل أو الأيقونات **Pre iconography**، التي تصف المعنى الأولي أو الطبيعي للصور، وتنقسم أيضاً إلى الجوانب الحقائقية **factual**، والتعبيرية **rexpressional** كما هو الحال على سبيل المثال، بالنسبة لصورة المدينة الصاخبة، إذ

المدينة هي الجانب الحقائقي، وطابعها الصاخب هو الجانب التعبيري. ويتطلب وصف صور ما قبل صناعة الأيقونات خبرة عملية؛ فالمدينة والمشهد الصاخب على سبيل المثال، من السهل التحقق منهما، بالنسبة لأولئك الذين شاهدوا مثل هذه الكيانات من قبل. أما الطبقة الثانية فهي صناعة الأيقونات، التي تحدد معالم المعنى الثانوي، أو التقليدي الذي يتم تصويره في الصورة (كصورة روما في مقابل أي مدينة أخرى على سبيل المثال). ويتطلب التحقق التماثيلي الإحاطة بالمصادر الأدبية، كالتآلف على سبيل المثال، مع ملامح روما كمدينة. أما الطبقة الثالثة فهي دراسة الأيقونات أو الرموز الفنية **iconology**، التي تمثل المعنى الجوهري أو القيمة الرمزية للصورة. ويتطلب التحقق من المعنى الجوهري للصور وتفسيره، الذي قلما يكون معروفا للفنان، وربما يكون مختلفا بشكل قاطع عن ذلك الذي قصد الفنان التعبير عنه بوعي- يتطلب "حدسا تحليليا تركيبيا **synthetic** مرتبطا بالسيكولوجية الشخصية" **(Panofsky, 1955, p. 41)**. ومثل هذا المعنى "يتم إدراكه بالتحقق من تلك الأسس التي تكشف عن الاتجاه الأساس لأمة ما، أو عصر، أو طبقة، أو ديانة، أو عقيدة فلسفية ما" **(Panofsky, 1955, p. 30)**. فبالنسبة للبعض، على سبيل المثال، يمكن لروما أن تعني مدينة ذات تاريخ ثري، بينما يمكن أن تعني لآخرين المكان الذي تقع فيه الفاتيكان، والمعنى الجوهري للصور، كما يبدو هو الأصعب في تفسيره. ومن الممكن للتفسير أن يكون مغرقا في الذاتية، إذ يتوقف على خبرة المرء وخلفيته.

وبعد نصف قرن من بانوفسكي **Panofsky** تقريبا، قسَّم إيكنز وجراهام **Eakins and Graham (1999)** خصائص الصور إلى ثلاثة مستويات من التعقد المتزايد، ففي المستوى الأول تكون الخصائص ملامح بدائية كاللون، والنسيج، والشكل. ومثال استفسارات الاسترجاع القائمة على خصائص صور المستوى الأول، المثلث الأحمر **red triangle**، حيث "الأحمر" يدل على اللون، و يوحي

"المثلث" بالشكل. وخصائص صور المستوى الثاني ملامح منطقية، تتصل بهوية الأشياء المسجلة في الصورة. ويمكن للأشياء أن تكون عامة (كمبنى المكتب على سبيل المثال) أو محددة (كبرج ويليز **Willis Tower**، الذي كان يعرف ببرج سيرز **Sears Tower**، في شيكاغو، على سبيل المثال). وخصائص صور المستوى الثاني أكثر تعقدا من خصائص المستوى الأول؛ إذ يدل الأول منطقيا على هوية الموضوعات في صورة ما، أو يشتمل على معلومات تتجاوز حدود المستوى البدائي، فموضوع الصورة على سبيل المثال مبنى إداري، وليس شكلا أو نسيجا بعينه. وخصائص صور المستوى الثالث مجردة، وتنطوي على قدر لا يستهان به من الحجج أو البراهين، حول معنى وهدف الموضوعات أو المناظر الواردة في صورة ما **(Eakins & Graham, 1999)**. وتتصل هذه الخصائص "بحيثيات **aboutness**" أو محتوى أو مضمون الصور، كالجو الاحتفالي المصور في منظر حفل على سبيل المثال.

ومن الممكن اكتشاف بعض مظاهر التشابه بين طبقات بانوفسكي **Panofsky** الثلاث للمعنى بالنسبة للصور، وخصائص الصور ثلاثية المستويات التي اقترحها إيكنز وجراهام **Eakins and Graham (1999)**. وتتعلق الطبقة الأولى في كلتا الحالتين "بالخواص **ofness**" أو الخصائص المادية للصور، إلا أن بانوفسكي أضاف الجانب التعبيري للصور في هذا المستوى، ولم يحلل الصور إلى بكسلات(*) **Pixels**، وهي المقياس الذي نشأ فيما بعد بالنسبة للصور الرقمية. وتغطي الطبقتان الثانية والثالثة في كلا الخطتين الحيثيات أو الخصائص الموضوعية للصور، المتعلقة بمحتوياتها الدلالية. ومن الممكن التحقق من بعض الخصائص الموضوعية بموضوعية (كتمييز روما عن غيرها من المدن على سبيل المثال)، بينما يمكن التأكد من البعض الآخر ذاتيا فقط (مثل روما بوصفها مدينة تاريخية أو مركزا دينيا). وأيا كانت الطريقة التي تقسم

(*) أدق دقائق النقاط التي تكون الصورة. (المترجم)

بها خصائص الصور إلى فئات، فإنها تشكل الأسس بالنسبة لاسترجاع معلومات الصور الثابتة، في كل من نهجي استرجاع الصور اللذين نناقشهما فيما بعد.

استرجاع الصور الثابتة القائم على الوصف:

يعتمد النهج القائم على الوصف لاسترجاع الصور الثابتة، الذي يعرف أيضا بالنهج القائم على ما وراء البيانات، أو القائم على المفهوم، أو القائم على النص - يعتمد على التوصيفات التي يوفرها اختصاصيو المعلومات. فمن الممكن وصف الصور الثابتة بسماتها أو خصائصها المادية (كصورة الدببة على سبيل المثال)، بالإضافة إلى حيثياتها أو خصائصها الموضوعية (كالحيوانات المهددة بالانقراض مثلا). وغالبا ما تسمى مثل هذه التوصيفات بما وراء البيانات **metadata** في أيامنا هذه؛ وقد أسمى ترانت **Trant (2004)** نهج استرجاع معلومات الصور هذا، بالقائم على ما وراء البيانات. وإذا ما كانت التوصيفات تتعلق بالحيثيات فقط، فإننا يمكن أن نسمي نهج استرجاع الصور بالقائم على المفهوم **Concept-based (Rasmussen, 1997)**. وإذا ما كانت التوصيفات معدة بالنصوص المصاحبة للصور، فإننا يمكن أن نسمي نهج استرجاع الصور بالقائم على النص **text-based (Goodrum, 2000)**، وحينئذ تصبح هذه التوصيفات الأسس بالنسبة لاسترجاع الصور الثابتة.

وتستخدم كل من الكلمات المفتاحية واللغات المقيدة، في استرجاع الصور الثابتة، وغالبا ما تشتق الكلمات المفتاحية واللغات المقيدة، في استرجاع الصور الثابتة، وغالباً ما تشتق الكلمات المفتاحية من العناوين الفرعية، أو المعلومات النصية الأخرى المصاحبة للصور الثابتة (كمقدمة فهرس أو دليل معرض الصور الضوئية على سبيل المثال). كذلك يمكن للمتعاملين اختيار مصطلحات الاستفسارات نفسها حين الاسترجاع. وتشمل اللغات المقيدة التي تستخدم في استرجاع الصور مكنز الفنون والعمارة **Art and Architecture Thesaurus**، وخطة تصنيف الفنون

والأيقونات **ICONCLASS**، ومكنز المواد التصويرية **Thesaurus for Graphic Materials**. ويمكن العثور على المزيد من الوصف لهذه اللغات المقيدة في راسموسن **Rasmussen (1997)**. وفي خطوة أكثر تقدما حلل جورجنسن **Jorgensen (2003)** خصائص هذه اللغات المقيدة (الموضوعات والبشر على سبيل المثال)، من حيث توزيع تكرارها، لمساعدة المتعاملين على الإحاطة بتطبيقها على نحو أفضل. كذلك تمت تجربة المكانز المرئية التي تستعمل البدائل المرئية، بالإضافة إلى التعبيرات اللفظية، لأغراض استرجاع الصور، في بعض الحالات **(Mostafa, 1994; Rasmussen, 1997)**. وتبدو هذه الطريقة مماثلة للاستفسار بالمثال، وهو أسلوب للاسترجاع، نوقش بإيجاز في القسم ٥/ ١/ ٢/ ٣ الخاص بتوسعة الاستفسارات.

وبالإضافة إلى استرجاع الصور بناء على حيثياتها، أو خصائصها الموضوعية، هناك الكثير من الخصائص الذاتية للوصول إلى الصور الثابتة، باتباع النهج القائم على الوصف. والتاريخ، والحجم، ونوع الصورة مجرد أمثلة قليلة. ولا تختلف عمليات البحث من هذا النوع كثيرا، عن البحث بالحقل في استرجاع المعلومات النصية. ولا تزال التقنيات التي تطورت لاسترجاع المعلومات النصية (كالبحث البوليني، والبتر على سبيل المثال) قابلة للتطبيق في استرجاع الصور الثابتة، مع النهج القائم على الوصف. ويكمن الاختلاف الرئيس بين استرجاع النصوص واسترجاع الصور، في تعريف الحقول أو تحديد معالمها؛ فالحقول في استرجاع الصور الثابتة، يتم تحديد معالمها تبعا لخصائص الصور لا وفقا لخصائص المعلومات النصية.

ومع ذلك فإن نهج استرجاع الصور القائم على الوصف، يرث جميع المشكلات المرتبطة بالتدخل البشري في العملية، كعدم الاطراد والتكلفة المرتفعة. يضاف إلى ذلك أنه لا مفر من الذاتية **subjectivity** في تفسير حيثيات الصور، وعلى درجة أكثر ارتفاعا مما هي عليه الحال في استرجاع النصوص؛ نظرا لأن تفسير حيثيات الصور أو محتواها

الموضوعي يتوقف بشكل كبير، على خبرة المرء الشخصية وخلفيته؛ فمن الممكن على سبيل المثال، لكأس النبيذ في صورة ما، أن يكون دليلا على الاحتفال بالنسبة لشخص ما، بينما هو رمز السكر بالنسبة لآخر. وقد لا تشتمل الصور نفسها على مفاتيح صريحة للتفسير. وقد دخل النهج القائم على المحتوى إلى مجال استرجاع الصور، جزئيا، للتغلب على مشكلة اتباع النهج القائم على الوصف هذه. وتشمل العوامل الأخرى التي أسهمت في ظهور النهج القائم على المحتوى، تطور التقنيات، ونمو معلومات الصور في العصر الرقمي.

استرجاع الصور الثابتة القائم على المحتوى:

يتم اتباع النهج القائم على المحتوى لاسترجاع الصور الثابتة، بمعالجة خصائص الصور المحددة في الاستفسارات آلى. وهذه الخصائص مشتقة من محتويات الصورة، ومن هنا سمي هذا النهج بالنهج القائم على المحتوى. وكما بينا في ٩/ ٢/ ١ فإنه يمكن تقسيم خصائص الصور إلى مستويين أو طبقتين مختلفتين **(Eakins & Graham, 1999; Panofsky, 1955)**. إلا أن نظم استرجاع الصور باتباع النهج القائم على المحتوى، لا تعمل في الوقت الراهن بأقصى فاعلية، إلا في المستوى الأول فقط، أي الملامح البدائية للصور كاللون، والنسيج، والشكل **(Eakins & Graham, 1999)**.

والمضاهاة هي التقنية الأساس في أي عملية للاسترجاع، واسترجاع الصور الثابتة بالنهج القائم على المحتوى ليس استثناء؛ فخصائص الصور المحددة في الاستفسار ينبغي أن تضاهي تلك الخصائص الممثلة في نظام استرجاع الصور، لكي يكون من الممكن الحصول على أي نتائج. وقد تم شرح المضاهاة الأساس، أو تقنية الاسترجاع بالنسبة للصور الثابتة بناء على اللون، والنسيج، والشكل، بوضوح واستفاضة في إيكنز وجراهام **Eakins and Graham (1999)**، وإن كان من الممكن لتطبيقات بعينها أن تنطوي على بعض الاختلافات.

ويستند استرجاع الصور الثابتة باللون إلى الرسوم البيانية **histogram** للألوان؛ إذ يتم تحليل كل صورة يشملها النظام، لحساب الرسم البياني للون، الذي يبين نسبة بكسلات **pixels** كل لون في الصورة، ثم يتم بعد ذلك اختزان الرسم البياني لألوان كل صورة في النظام للاستخدام في المستقبل. وبإمكان المتعامل، أثناء البحث، إما تحديد النسبة المرغوبة لكل لون (٣٠ بالمئة أبيض، و٣٠ بالمئة أزرق، و٤٠ بالمئة برتقالي، على سبيل المثال) وإما تقديم صورة كمثال (أي الاستفسار بالمثال) يتم منها حساب رسم بياني آخر للألوان. وبأي من الطريقتين تتم المضاهاة لاسترجاع تلك الصور، التي تضاهي الرسوم البيانية لألوانها تلك الرسوم البيانية للاستفسار إلى أقصى حد. ومن الممكن للنتائج التي يتم الحصول عليها، من بعض نظم استرجاع الصور باللون، أن تبدو مثيرة للإعجاب فعلاً **(Eakins & Graham, 1999)**.

ويستند استرجاع الصور الثابتة بالنسيج إلى الخصائص النصية **textual**، كالتباين **contrast** والاتجاهية **directionality**. وقد اتبعت أساليب متنوعة لقياس التشابه النصي بين الاستفسارات والصور المختزنة. فبادئ ذي بدء، تحسب هذه الأساليب للمعان **brightness** النسبي لثنائيات بكسلات **pairs of pixels** منتقاة من كل صورة، ثم يتم بعد ذلك حساب مقاييس أخرى لنسيج الصورة، كدرجة التباين، ومدى الخشونة، والاتجاهية، والانتظام (أو الدورية) والعشوائية. ومن الممكن صياغة استفسارات النسيج بطريقة مماثلة لتلك التي تصاغ بها استفسارات اللون، باختيار أمثلة من الأنسجة المرغوبة من لوحة الألوان **palette**، أو بتوفير صورة استفسار كمثال. ويسترجع النظام صورا بمقاييس نسيج مماثلة في قيمتها إلى أبعد حد للاستفسار. واسترجاع الصور بالنسيج مفيد بوجه خاص في التمييز بين مناطق الصور ذات الألوان المتشابهة (كالسماء والبحر، أو الأوراق والحشائش على سبيل المثال).

وقد تناول ستكس Stix (2006) بالوصف، تقنية ترتبط ارتباطا وثيقا باسترجاع الصور الثابتة بالنسيج؛ إذ يتم أولا تحليل كل صورة في نظام استرجاع معلومات الصور، بواسطة الحاسبات، للتحقق من الخصائص المميزة التي تتكون من مناطق داكنة **dark** محاطة بمناطق ناصعة **light**، أو العكس، ثم يتم بعد ذلك تجميع بعض الخصائص معا في مجموعات ثلاثية، بناء على حساب المسافة الفاصلة بين كل خاصية وأخرى. وتمثل كل خاصية النقطة المركزية وهي بقعة مربعة مكونة من ١٠ × ١٠ بكسلات. ويسمى تجميع ثلاث بقع، ثلاثية **triplet**.ويتم اختزان مكان كل ثلاثية كمدخل في جدول ضخم، مصمم للحد قدر الإمكان، من كم العمليات الحسابية اللازمة للبحث عن أي مدخل بعينه. كذلك يتم تقسيم صورة الاستفسار إلى ثلاثيات، بطريقة مماثلة لتلك الخاصة بتجهيز الصور المستهدفة. ويتم في النهاية مضاهاة ثلاثيات صورة الاستفسار مقابل تلك الثلاثيات الموجودة في الجدول، في أثناء عملية الاسترجاع.

ويستند استرجاع الصور الثابتة بالشكل، إلى شكل الشيء أو الموضوع، إلا أنه مستقل عن الحجم أو الاتجاه؛ إذ يتم حساب عدد من الخصائص المرتبطة بشكل الموضوع، لكل موضوع يتم التحقق منه، في كل وثيقة مختزنة. وهناك نوعان رئيسان لخصائص الشكل، عادة ما يستخدمان: (١) الخصائص العامة أو الشاملة **global features**، كنسبة العرض إلى الارتفاع والدائرية، (٢) الخصائص الموضعية **local**، كمجموعات القطاعات الحدودية المتوالية. والاستفسارات التي تقدم لنظم الاسترجاع بالشكل، إما أن تكون نماذج صور أو مخططات يرسمها المتعامل مع النظام. ويتم الرد حينئذ على الاستفسارات، بحساب مجموعة الخصائص نفسها الخاصة بصورة الاستفسار، واسترجاع تلك الصور المختزنة التي تضاهي خصائصها، تلك الخصائص الخاصة بالاستفسار إلى أبعد حد.

ويبدو النهج القائم على المحتوى، باللون والنسيج والشكل، قادرا على تقديم نتائج مفيدة، على الرغم من أن بعض الخصائص قد أثبتت أنها أكثر فاعلية من غيرها بكثير (**Eakins & Graham, 1999**). ومن المسلم به بوجه عام، أن استرجاع اللون واسترجاع الملمس، يسفران عن نتائج أفضل من مضاهاة الشكل (**Faloutsos, et al., 1994**). إلا أن الشكل مفهوم واضح المعالم بما فيه الكفاية، ومن السهل التعرف عليه في التحقق من الموضوعات الطبيعية الكامنة في الصور. ويكمن جزء من المشكلة بالنسبة لمضاهاة الشكل، في صعوبة التمييز آليا بين أشكال الصدارة **foreground**، وتفصيلات الخلفية **background** في الصورة الطبيعية. وقد طورت جماعة البحث في الاستفسار بمحتوى الصورة **Query By Image Content (QBIC)**، على سبيل المثال، بعض طرق التصدي لهذه المشكلة (**Flickner, et al., 1995**).

ويعد استرجاع الصور القائم على المحتوى **Content-based image retrieval (CBIR)**، على المستويات العليا أو الدلالية، على الرغم من عدم إمكان إجرائه على نحو مرضٍ في الوقت الراهن- مطلبا من جانب المتعاملين. وتتوقف الاختراقات في هذا المجال أيضا إلى حد بعيد، على التطورات التي تتحقق في بحوث الذكاء الاصطناعي، أي أنها يمكن أن تتم عندما تتمكن الحاسبات فعلا من محاكاة المهام الفكرية للعقول البشرية. ولا يزال استرجاع الصور، بناء على الخصائص الدلالية أو الموضوعية، بحاجة لتدخل البشر، الأمر الذي يدل على عدم إمكانية الاستغناء عن النهج القائم على الوصف في الوقت الراهن.

تكامل نهجي استرجاع الصور:

يتعامل النهج القائم على المحتوى مع خصائص الصور التي يمكن تجهيزها آليا، إلا أنه لا يعمل الآن بفاعلية، إلا مع الخصائص الأولية للصور، كاللون، والشكل، والملمس (**Eakins & Graham, 1999**). ولا يمكن معالجة المستويات

العليا لخصائص الصور، التي سبقت مناقشتها، بنجاح لأغراض الاسترجاع؛ نظرا لأن الحاسبات لا تزال غير قادرة على إنجاز التحليل الذكي للصور. يضاف إلى ذلك أن نهج استرجاع الصور القائم على المحتوى، أحيانا ما يفشل على نحو مؤسف **(Stix, 2006)**. وفي مقابل ذلك، يمكن للنهج القائم على الوصف، معالجة المستويات العليا أو الدلالية لخصائص الصور، بالتحقق من الصور، وما تتضمنه من موضوعات، ووصفها يدويا، من حيث ماهيتها وما تمثله. إلا أن مضاهاة الخاصية الأولية (كالشكل المعين لسقف مبنى ما، على سبيل المثال) التي يمكن أن تتم بيسر في استرجاع الصور القائم على المحتوى، يمكن أن تكون مهمة غير عملية، في النهج القائم على الوصف. ولكل نهج سياقه الواضح، وله أيضاً مظاهر قوته في استرجاع الصور **(Chu, 2001)**.

وينمو البحث في استرجاع الصور ويتسع، وخصوصا في المجال القائم على المحتوى. ويبدو هذا النهج القائم على المحتوى، وهو أحدث النهجين، مستخدما في مشروعات البحث، على نحو أكثر كثافة واتساعا من النهج القائم على الوصف **(Chu, 2001)**. إلا أن النهج القائم على الوصف يظل هو الخيار الأول، بالنسبة للنظم العاملة فعلا (**Graham, 1999**، على سبيل المثال).

وتزداد الحاجة إلى تكامل كلا النهجين وضوحا بمرور الوقت، وإلى أن يتم التوصل إلى وسائل للتجهيز الآلي للمحتوى الدلالي للصور **(Cawkll, 1992)**. وكان إيكنز وجراهام **(Eakins and Graham (1999**، اللذان يسميان النهج القائم على الوصف بطريقة الكلمات المفتاحية، يعتقدان أن استعمال الكلمات المفتاحية وخصائص الصور معا، يمكن أن يثبت جاذبيته. وقد سجلا عدة أسباب لاقتراحهما؛ أولها أن الكلمات المفتاحية يمكن أن تستعمل لالتقاط المحتوى الدلالي للصور، بينما يمكن لمضاهاة الخواص الأولية أن يكمل ذلك على نحو عملي، بالتحقق من خواص الصور (اللون أو الملمس على سبيل المثال) التي يصعب تسميتها باتباع نهج الوصف. ثانيا،

من الممكن إحراز أرصدة نقاط مرتفعة للتحقيق والاستدعاء، عندما يتم، على سبيل المثال، استعمال تشابه النص واللون معا، أرصدة أعلى مما يمكن إحرازه عند استعمال أي منهما مستقلا. ثالثا، يمكن للاسترجاع بمجموعة من الطرق، باتباع بنى أو نظم معرفية مختلفة، أن يكون أكثر فاعلية مما يمكن أن يتم بأي طريقة بمفردها (Ingwersen, 1996). ويسمى إنزر (2008) Enser هذا النهج المتكامل بالاسترجاع الدلالي لمعلومات الصور.

وقد حاول بعض الباحثين المنتمين إلى أوساط الاسترجاع القائم على المحتوى فعلا، دمج المعلومات النصية والدلالية مع الخصائص البصرية للصور، في بحوثهم، الأمر الذي يدل على بداية واعدة لتكامل النهجين (Chu, 2001). كذلك نشرت في السنوات الأخيرة، تقارير تربط بين نهجي البحث المختلفين (مثل Barnard, et al., 2003; Chang, et al., 1997b). وفضلا عن ذلك أشار إيكنز وجراهام Eakins and Graham (1999)، في دراستهما، إلى أن الحواجز بين أوساط البحث المختلفة في المجال، تتهاوى تدريجيا. ومن الممكن لتكامل النهجين أن يحقق وصولا أفضل وأكثر كفاءة للصور الثابتة، وكذلك الأنواع الأخرى من المعلومات متعددة الوسائط.

استرجاع الأصوات:

تشمل الأصوات، التي تسمى أيضا بالسمعيات، الموسيقى غير اللفظية والحديث اللفظي، بالإضافة إلى الأصوات الطبيعية أو الاصطناعية الأخرى (كتغريد الطيور وضوضاء المرور على سبيل المثال). والموسيقى والحديث هما النوعان الرئيسان للمعلومات التي يتم استكشافها في استرجاع الأصوات.

وقد أصبح استرجاع المعلومات الموسيقية (Music IR (MIR تدريجيا، مجالا متميزا للبحث؛ حيث نجد الآن ورش العمل، واللجان، والمؤتمرات المكرسة خصوصًا للموضوع (Downie, 2003). وكذلك اللغة، أحد المكونات الأساس لاسترجاع

المعلومات، فإن للموسيقى معنى مستتر. إلا أن الموسيقى ربما تكون أكثر تعقدا من اللغة؛ إذ أن أي أداء لها يضيف طبقة من التعقد (Lippincott, 2002). وللموسيقى دلالتها الخاصة، ويمكن للحن أن ينقل معنى قد لا تكون الكلمات قادرة على التعبير عنه. وكما هو الحال بالنسبة للصور، يمكن للمقطوعة الموسيقية أن تفسر على نحو مختلف، من جانب ذوي الخلفيات المختلفة، وفي المواقف المختلفة. وفضلا عن السمات الموسيقية كالملحن، والأسلوب، والمؤدي، التي تستخدم في استرجاع المعلومات الموسيقية القائم على الوصف، للموسيقى خواص فريدة (كالنغمة **pitch**، والتناغم أو التوافق **harmony**، والجرس **timbre** على سبيل المثال)، التي تستند إلى أي نوع من استرجاع المعلومات الموسيقية القائم على المحتوى، يتم تطبيقه. إلا أن استرجاع المعلومات الموسيقية يواجه تحديات ضخمة، في أي من نهجي الاسترجاع؛ نظرا لطبيعة المعلومات الموسيقية نفسها **(Byrd & Crawford, 2002; Downie, 2003).**

والحديث أو الخطاب **speech**، شكل للاتصال أكثر ثراء من النص التحريري أو المكتوب، من نواح عدة، نظرا للمكون الصوتي الإضافي. فمن الممكن على سبيل المثال لجرس **sound** الصوت **voice** أن يساعد في تحديد هوية المتحدث **(Oard, 2000)**. وغالبا ما يسمى الحديث أو الخطاب، بالوثائق الناطقة **spoken**؛ فنشرات الأخبار، ومحادثات المؤتمرات، والمحادثات الهاتفية، من أمثلة الوثائق الناطقة. ونظرا لعدم إمكان انتزاع زبد **skimmed** المعلومات الصوتية أو مسحها أو تدقيقها **scanned** أثناء عملية الاسترجاع، يتعين على الإنسان أن يستمع إليها من البداية إلى النهاية، عند محاولة العثور على مكان جزء بعينه. ويستغرق الاستماع إلى حديث أو خطاب ما، بوجه عام، وقتا أطول بكثير مما يستغرقه تدقيق النص أو مسحه. وتحتاج المعلومات الصوتية ولاشك، إلى طرق استرجاع تختلف عن طرق استرجاع النصوص.

استرجاع المعلومات الصوتية القائم على الوصف:

لقد كان النهج القائم على الوصف، هو الطريقة الوحيدة لاسترجاع المعلومات الصوتية، قبل ظهور النهج القائم على المحتوى. فبإمكان الإنسان البحث عن الموسيقى باللحن، أو الأسلوب، أو الأدوات المستخدمة، وربما أيضا بموضوع الموسيقى، على سبيل المثال (**Lesk, 2005**). ومن الممكن استرجاع معلومات الحديث أو الخطاب، من بين نقاط وصول أخرى، بالمتحدث، والمكان، والتاريخ، والوسائط التي سجل عليها الحديث، وموضوع الحديث. والمشكلات الرئيسة في اتباع النهج القائم على الوصف، لاسترجاع المعلومات الصوتية، هي الوقت والدقة. ونظرا لأن المعلومات الصوتية غير قابلة للتصفح؛ فإن الأمر يستنفد الوقت إلى أقصى حد، لإعداد توصيفات حول المعلومات الصوتية لأغراض الاسترجاع. وإذا كانت سمات المعلومات الصوتية تنقسم إلى فئتين، هما الذاتية **ofness** والموضوع **aboutness**، فإن موضوع المعلومات الصوتية، كما هو الحال بالنسبة لأي نوع آخر من المعلومات، من الصعب على الإنسان الحكم عليه أو الفصل فيه بدقة. ويبدو ذلك حقيقة واقعة على وجه الخصوص، بالنسبة للمعلومات الصوتية غير اللفظية، كالموسيقى التي قد لا تكون قد قصدت للتعبير اللفظي.

استرجاع المعلومات الصوتية القائم على المحتوى:

يستعار البحث حول استرجاع الموسيقى القائم على المحتوى، من ثلاثة مجالات؛ هي استرجاع المعلومات التقليدي، وعلم الموسيقى، وإدراك الموسيقى أو فهمها (**Lippincott, 2002**). وتشمل السمات المشتركة التي توضع في الحسبان، في استرجاع المعلومات الموسيقية القائم على المحتوى، النغمة، والمدى الزمني، والتناغم، والجرس. وتدل النغمة على وجه التحديد، على مدى ارتفاع الصوت أو انخفاضه. وتتعلق المدة بمدى استمرار النوتة الموسيقية **note**. أما التناغم فيعني نغمتين أو أكثر،

تسمعان في الوقت نفسه، بينما يعني الجرس نوعية النوتة الموسيقية (Byrd & Grawford, 2002; Downie, 2003). ويتم تحليل هذه السمات آليا لأجل التحقق من الموضوعات الموسيقية (Birmingham et al., 2002)، أو الوحدات المتفرقة من المعلومات اللحنية (Downie, 2000) التي تصبح الأساس بالنسبة لاسترجاع المعلومات الموسيقية القائم على المحتوى. فقد اقترح داوني Downie (2000) على سبيل المثال، استرجاع المعلومات الموسيقية بفترات س-جرام n-gram (أي المسافة ما بين النوتات الموسيقية) التي تشتمل عليها الألحان. وتشكل س-جرامات n-grams (راجع أيضا ٥/ ١/ ٢/ ١) وحدات متفرقة من المعلومات اللحنية، على النحو نفسه إلى حد بعيد، الذي تشكل به الكلمات الوحدات المتفرقة للغة. ومن ناحية أخرى تبين لبرمنجهام وآخرين Birmingham et al. (2002)، أن نماذج ماركوف[*] Markov models المستترة، عملت بشكل جيد في نظام استرجاع الموسيقى ميوزآرت Musart الذي قاموا بتطويره. والخوارزميات الخاصة باسترجاع المعلومات الموسيقية القائم على المحتوى، في الواقع أكثر عددا بكثير من الاثنتين اللتين ذكرتا توا. ويبدو أن هناك كثيرا من الخوارزميات الآلية لاسترجاع الموسيقى، بقدر ما هنالك من أعداد المتعاملين (Downie, 2000).

وفي نظام لاسترجاع المعلومات الموسيقية بناء على المحتوى، يتم تجهيز كل من معلومات مرصد البيانات والاستفسارات، باتباع خوارزمية معينة، لأغراض المضاهاة؛ فبإمكان المتعامل النهائي إدخال الاستفسار بالغناء أو الدندنة humming، أو باستخدام آلة موسيقية ما (Birmingham, Dannenberg & Pardo, 2006 على سبيل المثال). إلا أن البحث حول استرجاع الموسيقى، في مقابل البحث حول استرجاع المعلومات النصية، نادر، حتى وإن كانت هناك مؤتمرات سنوية (مثل

(*) العشوائية Stochastic. (المترجم)

(International Symposium on Music Information Retrieval تعقـد في السنوات الأخيرة، من أجل استرجاع الموسيقى على وجه الخصوص. ولا تزال النظم العاملة فعلا لاسترجاع المعلومات الموسيقية بناء على المحتوى نادرة الوجود.

وفي مقابل ذلك، يتضمن الإنتاج الفكري عدداً كبيرا جدا من البحـوث حـول اسـترجاع المعلومـات الـصوتية القـائم عـلى المحتـوى (مثـل **Goodrum, & Rasmussen, 2000; Spärck Jones, et al., 1996**)، كما أنشئ مسار لاسترجاع الوثائق الناطقة (مسار استرجاع الوثائق الناطقة **SDR track**) في سلـسلة مـؤتمرات اسـترجاع النـصوص **TREC**، مـن تـرِك -٦ حتـى تـرِك -٩ (**Voorhess & Garafolo, 2000**). والخطوة الأولى في استرجاع الخطاب هي فصل القطاعات غـير الصوتية عن القطاعات الصوتية. وتشمل القطاعات غير الصوتية أو المكابح الصوتية **speech constraints**، الوقائع غير الصوتية (كالتنفس المرتفع وطقطقة اللسان عـلى سبيل المثال)، ومظاهر تعثر النطق أو عيوب النطـق **disfluencies** (كـالكلمات غـير مكتملة النطق، وحالات التوقف أو الـتردد، والفأفـأة أو التهتهـة **hesitations** عـلى سـبيل المثـال)، والعنـاصر أو المكونـات الوظيفيـة (مثـل و**and** وبالإضـافة إلى **in addition**)، والموسيقى أو الضوضاء التي تغشى الصوت. وبعد إزالة هذه القطاعات غير الصوتية، يمكن إجراء المزيد من التحليل للقطاعات الصوتية تبعا لمختلف سمات الخطاب أو الحديث.

وتقسيم الحديث أو الخطاب أو تحليله إلى قطاعات، خطوة تثير التحديات، إلا أنه لا غنى عنها في استرجاع الخطاب. وبينما يمكـن نـسخ أو تحويـل التـدفق المـستمر لمعلومات الحديث مباشرة، دون أي تحليل مسبق، فـإن التحليـل يكفـل عـدة مزايـا، بالمقارنة بالحـل المـستقيم المبـاشر **straightforward**؛ إذ مـن الممكـن أولا اقتطـاف معلومات مهمة (كالفصل بين المتحدثين، والتحقق من هويات المتحدثين عـلى سـبيل

المثال) أثناء التحليل. ثانيا، يمكن للتحليل للمسبق أن يؤدي إلى تجنب المشكلات الناشئة عن الانقطاع اللغوي أو الفجوة اللغوية عندما يتغير المتحدث. ثالثا يحد استبعاد القطاعات غير الصوتية، وتقسيم البيانات إلى قطاعات أقل طولا، من وقت إجراء العمليات الحاسبية على نحو ملحوظ، كما ييسر فك الشفرات في مرحلة لاحقة **(Gauvain, Lamel & Adda, 2000)**.

وقد تم تطوير خوارزميات عدة، للتعرف على الحديث وفهم الحديث. والتعرف على من يتحدث مشكلة مختلفة قليلا عن التعرف على ما يقال **(Lesk, 2005)**. ويتوقف الإلمام بالحديث على فهم الحديث، الذي يعتمد أكثر على معالجة اللغة الطبيعية **(NLP)**، وهو موضوع نناقشه في الفصل الثاني عشر. فالنغمة، والسرعة، والكثافة (أي التوقف بين الجمل أطول من التوقف بين الكلمات) من بين السمات التي عادة ما تستخدم لاسترجاع الأصوات. وفضلا عن ذلك يستخدم نوع المتحدث، وإحصاءات استعمال الكلمات، والكلمات المفتاحية التي تميز كل فئة، وذلك بالارتباط مع مفردات اللغة التي تنشأ على وجه التحديد من أجل الإلمام بالحدث وفهمه **(Kubala, et al., 2000; The SRI MAESTRO Team, 2000; Wactlar, et al., 2000)**. كذلك تستخدم أساليب الاسترجاع التقليدية، كالوزن أيضا في استرجاع الحديث **(Spärck Jones, et al., 1996)**. إذ يتم تحويل الحديث آليا، بإنشاء النص المقابل الذي سيستخدم للمضاهاة أثناء عملية الاسترجاع. ومن الممكن تقديم الاستفسارات في شكل حديث، إذا ما توافرت تقنية للتعرف على الحديث لدى الطرف الخاص بالبحث.

وكما ذهب لسك **Lesk (2005)**، فإن هناك أربعة عوامل أو محددات تؤدي بقوة إلى صعوبة التعرف على الحديث:

١. استمرارية الحديث: فالحديث المتواصل أو المستمر أصعب في فهمه من الحديث الذي يتم فيه التوقف بين الكلمات والجمل.

٢. عدد المتحدثين: إذا كان الأمر يتطلب التعرف على جميع المتحدثين المشاركين في الحديث، فإن المهمة يمكن أن تكون أصعب مما هي عليه في حالة ما إذا كان متحدث واحد فقط هو موضع التحليل.

٣. محتوى الحديث: يمكن لمحتوى الحديث أن يتراوح بين أي شيء بلا موضوع محدد، وموضوع معين بمفردات لغوية محددة على نحو جيد، كما أنه من الأيسر بمكان إنجاز التعرف على الحديث في الحالة الأخيرة، مما يمكن أن يتم في الحالة الأولى.

٤. بيئة الحديث: هناك احتمالات كثيرة بالنسبة لبيئة الحديث؛ فمن الممكن من ناحية، للمتحدث أن يستعمل مكبرا جيدا للصوت في غرفة هادئة (كحديث المؤتمر على سبيل المثال)، وعلى الجانب الآخر يتحدث المتحدث في شارع صاخب، بدون مكبر صوت جيد (كالمقابلات في المواقع على سبيل المثال). والبيئة الأولى هي المثالية بالنسبة للتعرف على الحديث، بينما الثانية ليست كذلك.

وبالإضافة إلى هذه العوامل الأربعة، يمكن للهجة المتحدث أن تضيف إلى تعقد التعرف على الحديث أيضا.

استرجاع الصور المتحركة:

يستند استرجاع الصور المتحركة، نظرا لتكويناتها، على استرجاع الصور الثابتة والأصوات، الذي نوقش توا؛ فمن الممكن للصور المتحركة التي تتكون من الصور الثابتة بالإضافة إلى السمة المتحركة، أن تكون مصحوبة بالصوت أو غير مصحوبة

بالصوت. وأشرطة الفيديو، والأفلام السينمائية، والبرامج التلفزيونية، من أمثلة الصور المتحركة المصحوبة بالصوت، بينما تمثل الرسوم المتحركة أو الأفلام الصامتة، الصور المتحركة بدون صوت.

وقد اتبع النهج القائم على الوصف فى استرجاع الصور المتحركة. وتشمل التوصيفات الخاصة بالصور المتمركزة جوانب مثل المخرج السينمائي، وسنة الإنتاج، تلك الجوانب التي عادة ما يتم اختيارها للتعبير عن الخواص الذاتية والموضوعات. إلا أن استرجاع الصور المتحركة باتباع هذا النهج، عرضة للمشكلات نفسها (كالتكلفة المرتفعة والكفاءة المنخفضة على سبيل المثال) التي غالباً ما تكتنف استرجاع الأنواع الأخرى من المعلومات متعددة الوسائط.

وينطوي استرجاع الصور المتحركة، باتباع النهج القائم على المحتوى، بالضرورة، على ثلاث مهام رئيسة، وإن كان من الممكن إجراء الاسترجاع ببعض الأشكال المختلفة. والمهمة الأولى هي عزل المكون الصوتي، إذا كان هناك مكون صوتي، عن الصورة المتحركة، وتجهيز المعلومات الصوتية. والمهمة الثانية هي تقسيم الصور المتحركة إلى صور ثابتة وفقاً للقطات أو أي سمات أخرى. والمهمة الثالثة هي تجهيز الصور الثابتة. وبعبارة أخرى، فإنه من الممكن لاسترجاع الصور المتحركة، أن يختزل إلى استرجاع الصوت والصورة الثابتة. ويكمن ما يميز استرجاع الصور المتحركة فعلاً، عن استرجاع الصوت والصور الثابتة، في تجزئة الصور المتحركة. ونظراً لأننا قد تناولنا استرجاع الأصوات واسترجاع الصور الثابتة في الأقسام السابقة؛ فسوف تركز هذه المناقشة على تجزئة الصور المتحركة لأغراض الاسترجاع.

فالصور المتحركة ينبغي أن تجزأ أو تفكك، لكي تصبح قابلة للاسترجاع. وعادة ما تتخذ المناظر أو الوقائع وحدات دلالية، من أجل تجزئة الصور المتحركة. ويتم التحقق منها باكتشاف التغير في هذه العوامل، كتوزيع اللون، والملمس، والحزوز

cuts، والظلال **fades**، والتلاشي، وومضات الكاميرا، آليا. ويتكون المنظر بدوره من عدة لقطات، يمكن تحليلها أيضاً إلى عدة أطر لكل لقطة. ومن ثم فإن التقسيم يمكن أن يتم وفقاً للمنظر، أو اللقطة، أو الإطار.

ويعرف تجزىء الصور المتحركة باللقطة، باكتشاف أطر اللقطات **shot boundary frames detection (SBD)**. وتقوم معظم أساليب اكتشاف حدود اللقطة بحساب التشابه بين الأطر المتجاورة، وتمييز **flag** التغير المحتمل للقطة، عندما ينخفض ذلك التشابه إلى ما دون حد معين **(Smeaton, 2004)**. وتعمل هذه الأساليب بشكل مناسب إلى حد معقول، إذا كان بالصور المتحركة حدود لقطات واضحة، بحيث يرد الإطار الأول للقطة الجديدة بعد الإطار الأخير من اللقطة السابقة مباشرة. إلا أن كثيراً من الصور المتحركة تنطوي على تحول تدريجي من لقطة إلى أخرى (كحالات التلاشي الداخلي وحالات التلاشي الخارجي، وحالات الذوبان، والمظاهر الخارجية **morphs**، والرتوش أو اللمسات **wipes** على سبيل المثال، لجعلها أكثر متعة من الناحية الجمالية. إنها مثل هذه المؤثرات التجميلية في الصور المتحركة، التي تتحدى سهولة اكتشاف حدود اللقطات. وفي مقابل ذلك، تميل بعض أساليب اكتشاف حدود اللقطات (كاكتشاف الحواف على سبيل المثال) لأن تتعامل مع التحولات التدريجية بشكل مناسب، إلا أنها من المرجح أن تسفر عن صور موجبة خادعة في اللقطات الخافتة أو البعيدة عن التركيز **(Smeaton, 2004)**.

وكما سبق أن لاحظنا، فإن اللقطة تتكون من أطر مجمعة معا، بناء على بعض السمات المشتركة. ويسمى الإطار الذي يمثل كل لقطة الإطار الرئيس **keyframe**. ويعني اختيار إطار رئيس، العثور على الإطار الذي يمثل المجموعة المتوسطة، من الأطر في اللقطة على خير وجه. ويمكن للإطار الرئيس أن يكون في بداية اللقطة، أو فى نهايتها، أو في أي موضع متوسط، إلا أنه لم يتم حتى الآن إجراء دراسات لتحديد

أي أساليب اختيار الإطار الرئيس أكثر نجاحاً في التطبيق (Smeaton, 2004)، غير أنه غالباً ما ينظر إلى اختيار الإطار الرئيس بوصفه "مهمة مغلقة black art"، وتنطوي أكثر الطرق شيوعاً على اختيار الإطار الرئيس من منتصف اللقطة.

ومن الممكن بعد ذلك، وضع مجموعة من الأطر الرئيسة معاً لتشكل القصة المصورة **storyboard** للصور الثابتة الموضحة أو الشارحة، لتمثيل كل مشهد. وعلى الرغم من أن التجميع الآلي للقطات في مشاهد منطقية، عن طريق الأطر الرئيسة، يحظى الآن بقدر كبير من الاهتمام، فقد أثبتت الجهود من هذا النوع صعوبتها، إلا في المجالات المنضبطة تنظيمياً للصور المتحركة (كنشرات الأخبار التلفزيونية، التي عادة ما تنطوي على الأفراد العاملين بالربط، والأرصدة التمهيدية، وسلسلة القصص المتفرقة). وبالإضافة إلى طريقة الإطار الرئيس، يستخدم أسلوب انتزاع زبد الفيديو **skimming** لإنتاج سلسلة من مقاطع الفيديو القصيرة، التي يتضمن كل منها التفصيلات الأساسية لتتابع واحد، أو سلسلة من اللقطات ذات الدلالة اللغوية **(Eakins & Graham, 1999)**. وتشكل تلك الكيانات، الأساس بالنسبة لاسترجاع الصور المتحركة. وعادة ما يوفر تتابع الصور المتحركة معلومات أكثر مما توفرها الأطر الرئيسة، لأغراض الاسترجاع. ومن الممكن للصور المتحركة نفسها، أن تكون بمثابة استفسارات، إذا ما كانت مقاطع الفيديو القصيرة مقتطفة، ومتاحة في نظام استرجاع المعلومات. فالاستفسار بالمثال المتحرك من الحالات الجديرة بالاهتمام **(Chang et al., 1997a)**.

وهناك نهج آخر، ربما يكون الأبسط والأكثر شيوعاً، وهو مضاهاة نص استفسار المستفيد، مقابل تسجيل خطي من المحتوى الناطق المقتطف من الصورة المتحركة. وفي استرجاع الوثائق الناطقة، يمكن لهذا الأسلوب أن يظل فعالاً إلى حد بعيد، حتى وإن كان معدل الخطأ في الكلمات في حدود ٥٠ بالمئة **(Spärck Jones et**

(al., 1996. ومن بين مضامين نتيجة عملية البحث هذه، أن يبدو استرجاع الصور المتحركة بناء على الحديث الذي يتم التعرف عليه واعداً. وأفضل مثال لنظام استرجاع الفيديو الذي يدعم البحث بالنسخة الخطية، أو البحث عن طريق العبارات الجاذبة المغلقة، هو علامة الطريق، التي تطورت في مشروع مكتبة الفيديو الرقمي إنفورميديا (Informedia Digital Video Library (www.informedia.cs.cmu.edu بجامعة كارينجي ملون Carnegie Millon University (Hauptmanm & Witbrock, 1997). وأحياناً ما يكون من الممكن استعمال المعلومات الصوتية التي تتضمنها الصور المتحركة، للتخلص من الغموض في تجزيء وتفسير الصور المتحركة. والعكس صحيح أيضاً (Wactlar, et al., 2000).

وربما يتبين لنا من هذه المعالجة الموجزة، أن التجزيء الآلي للصور المتحركة، ليس بالمهمة اليسيرة؛ فالتجزيء عملية معقدة متعددة الخطوات. فالفيديو على سبيل المثال، يمكن أن يتكون أولاً من القصة، ثم من الأحداث، واللقطة والإطار (Boykin & Merlino, 2000; Smeaton, 2004). وبينما يبدو التجزيء باللقطة والإطار مباشراً، فإن التجزيء بالقصة والأحداث يتطلب جهوداً بشرية، أو ذكاء اصطناعياً. وهذا مبرر آخر لضرورة الجمع بين النهج القائم على المحتوى والنهج القائم على الوصف، في استرجاع الوسائط المتعددة، في ظل الوضع الراهن لتقنيات الاسترجاع (Wan & Liu, 2008).

استرجاع الوسائط المتعددة على الإنترنت:

اشتهرت عنكبوتية الإنترنت بقدرتها على تداول المعلومات متعددة الوسائط، ومن ثم فإن عدداً كبيراً من نظم استرجاع الإنترنت (جوجل على سبيل المثال) يدعم البحث في الوسائط المتعددة، سواء كانت هذه النظم وحدها، أو بالتعاون مع

الشركات المتخصصة في إدارة معلومات الوسائط المتعددة، على نحو يكفل للمستفيد النهائي الوصول إلى ثروة الوسائط المتعددة على العنكبوتية:

وهناك ثلاث طرق مختلفة قائمة على الوصف، في متناول المستفيدين النهائيين، تسترجع معلومات الوسائط المتعددة من العنكبوتية:

١. بالكلمات المفتاحية: إذ يستطيع المستفيدون استرجاع معلومات الوسائط المتعددة بالكلمات المفتاحية، تماما كما يفعلون في بحث المعلومات النصية. وتعمل هذه الطريقة بكفاءة على وجه الخصوص، في المواقع المخصصة لعمليات البحث عن الصور، مثل بكسيرش **Picsearch (www.picseareh.com)**.

٢. بنوع الوسائط المتعددة: فالوسائط المتعددة اسم جامع لمختلف أنواع المعلومات غير النصية، كالصور، وتسجيلات إم بي ثري **MP3**، والفيديو. ويمكن لأنواع الوسائط المتعددة هذه، أن تكون بمثابة نقاط وصول في عملية الاسترجاع. وتستخدم بعض نظم استرجاع الإنترنت (مثل ألتافيستا **Altavista**) أزرار المذياع، أو تقنية أخرى مماثلة، ليختار بها المستفيديون النوع الذي يودون استرجاعه من الوسائط المتعددة.

٣. بامتداد اسم الملف: فعادة ما تكون لأسماء الملفات في البيئة الرقمية، ما يصل إلى ثلاثة أحرف في امتداد ملفها، لتحديد نوع الملف. وفي حالة ملفات الوسائط المتعددة، فإن **gif** و**Jpg**، على سبيل المثال امتدادات عادية للملفات، بالنسبة للصور الثابتة. وعادة ما تنتهي أسماء الملفات، الناطقة بـ **au**، أو **wav**، أو **mp3**. أو ما شابه ذلك، وتشمل امتدادات أسماء الملفات بالنسبة للصور المتحركة **mpg** و**mov**. ومن الممكن تضمين امتدادات أسماء الملفات هذه في

الاستفسارات، وذلك لاسترجاع نوع معلومات الوسائط المتعددة المحدد (مثل greatwall.gif كاستفسار للبحث).

والنهج القائم على المحتوى متاح أيضاً على الإنترنت، لاسترجاع معلومات الوسائط المتعددة. وتدعم بعض نظم استرجاع الإنترنت، كلا النهجين؛ فبإمكان المتعاملين، على سبيل المثال، بدء عملية بحث ما بنهج الكلمات المفتاحية القائم على الوصف، ثم تتم توسعة البحث بالسمات القائمة على المحتوى، كاللون والشكل على سبيل المثال. ويكفل بحث الصور في **Bing.com** للمتعاملين، الحصول على المزيد من الصور المماثلة للصورة المسترجعة، باستعمال الكلمات المفتاحية، وذلك بالنقر ببساطة على رابطة "المزيد من الصور المماثلة **more similar images**" **(Schwartz, 2008)**. والنظم الأخرى مصممة للتعامل فقط مع استرجاع معلومات الوسائط المتعددة، القائم على المحتوى. إلا أن النظم من هذا النوع (مثل **Convera's Visual Retrievalware**) تتاح للمتعاملين مع الإنترنت، عندما تكون النظم في مرحلة الاختبار التجريبي **beta test** فقط. وبمجرد انتهاء الاختبار، تصبح هذه النظم تعمل بمقابل رسوم، ولا تتاح في النطاق العام. ويمكن لنظم أخرى أن تتوقف عن العمل نظراً للافتقار إلى الترويج أو الإدارة المناسبة.

واسترجاع معلومات الوسائط المتعددة على الإنترنت، في سبيله لأن يصبح أيسر مما كان عليه من قبل. ويرجع الفضل في ذلك إلى التطورات التقنية، وجهود الباحثين. إلا أن استرجاع الوسائط المتعددة بوجه عام، لا يزال أمامه شوط طويل، حتى يحقق الأداء المرضي. وقد لخص لسك **Lesk (2005)** يوما ما الوضع الراهن للمعرفة في استرجاع الوسائط المتعددة على النحو التالي:

بإمكاننا تصنيف الصور **pictures** على وجه التقريب إلى حد بعيد، وفقاً للون والملمس والشكل، ولا يكفي ذلك للتعامل مع مجموعات ملايين الصور الضوئية. وبإمكاننا تحقيق التعرف على الصوت، بطريقة تكفي تماما لأن يكون هذا التعرف مفيداً إلى حد ما، في تحليل التسجيلات الصوتية، ولدينا تقنيات محددة لبحث الموسيقى. ولدينا حيل متعددة لضغط الفيديو وتصفحه، إلا أنه ليس لدينا سوى تقنيات محدودة للبحث فيه **(p.115)**.

معلومات النصوص الفائقة والوسائط الفائقة:

يعني صدر الكلمة أو البادئة **hyper** في **hypertext** "الامتداد إلى أبعاد أخرى أو تعدد الأبعاد" **(Cornejio, 1997)**. وبذلك المعنى يتكون النص الفائق **hypertext** من نقاط ارتكازية **nodes** من المعلومات، وروابط بينها، في فضاء متعدد الأبعاد. والوسائط الفائقة **hypermedia** كذلك أيضاً، من حيث المفهوم والبناء، إلا أن الوسائط المتعددة لا تشمل النصوص فحسب، وإنما تشمل أيضاً معلومات الصوت والصورة.

بيد أن كلا من النصوص الفائقة والوسائط الفائقة، تختلف عن النصوص والوسائط المتعددة من النواحي التالية؛ فالنصوص الفائقة والوسائط الفائقة، تتخذ أولا شكل الشبكة، لا التكوين الخطي المسطح للنص والوسائط المتعددة. ثانياً، هناك الروابط التى تربط بين النصوص الفائقة والوسائط الفائقة مادياً، بينما الروابط بين الوثائق غير فائقة البنيان **nonhyperstructured** مجرد روابط مفاهيمية. ثالثاً، يتم وضع الروابط بين الوثائق فائقة البنيان **hyperstructured** من أجل المستفيد، بينما يتعين على المستفيد أن يعد الروابط بين الوثائق غير فائقة البنيان، حسبما تقتضي

الضرورة. وتفضي الخواص الفريدة للنصوص الفائقة والوسائط الفائقة، إلى طرق جديدة لاسترجاع المعلومات، في البنية فائقة التنظيم.

ويمكن للتصفح، لا البحث أن يكون النهج الرئيس، الذي يتبعه المتعاملون في استرجاع النصوص الفائقة والوسائط الفائقة؛ نظراً لأنه كما بينا في ٦/ ٢/ ٢، فإن البيئة فائقة البنيان تشجع أنشطة التصفح. وتتوافر نقاط الوصول أو مصطلحات الاستفسار كأسماء روابط فائقة، في عملية الاسترجاع هذه، بحيث تستبعد حاجة المتعامل لأن يختارها بنفسه. وقد يبدو اختيار مصطلحات البحث سهلاً بالنسبة لاختصاصي المعلومات المتمرس. إلا أنه بالنسبة للمستفيد النهائي مهمة معقدة، تتطلب الكثير، وتنطوي على تحليل للمفاهيم، والنظر في أوجه الاختلاف بين المصطلحات. يضاف إلى ذلك قدرة المستفيد على التحرك تلقائياً في فضاء استرجاع المعلومات متعدد الأبعاد هذا، بمجرد النقر على الروابط الفائقة، التي تربط بين مختلف النقاط الارتكازية للمعلومات. وعندما يقع الاختيار على البحث، كنهج للاسترجاع، فإن أسماء الروابط تعد أكثر أهمية من المصطلحات العادية، في الوزن والترتيب الطبقي، إذا ما توافرت تقنية الترتيب الطبقي. فكثير من نظم استرجاع الإنترنت، على سبيل المثال، تضع عناصر مسترجعة معينة، في مرتبة أعلى من غيرها، إذا ما كان مصطلح الاستفسار يضاهي اسم رابطة تتضمنه الوثيقة.

ومن ناحية أخرى، يكتنف استرجاع النصوص الفائقة والوسائط الفائقة، بالتصفح مشكلات محتملة، نظراً للبنيان الفائق. فمن الممكن أولاً، للمتعاملين أن يضلوا طريقهم في الفضاء الفائق، عند التصفح باتباع الروابط الفائقة ببساطة. ثانياً يمكن للمتعاملين أن يجدوا من الصعب إلقاء نظرة عامة على المعلومات التي يبحثون عنها؛ نظراً لأن انتباههم يمكن أن ينصرف بسهولة، نحو كثير جداً من الوثائق الأخرى المرتبطة. ثالثاً، يمكن للمتعاملين أن يجدوا صعوبة في العثور على معلومات

معينة، حتى وإن كانوا يعرفون أن هذه المعلومات موجودة فعلا في النظام. ويمكن لخرائط المواقع التي تنشأ في بعض مواقع العنكبوتية، أن تساعد في هذا الصدد. إلا أن هذه الخرائط لا تتوافر في جميع نظم استرجاع المعلومات فائقة البنيان، وإذا ما توافرت، فإنها قد لا تكون مفصلة كما ينبغي.

وييسر إدخال نهج البحث في البيئة فائقة البنيان، استرجاع النصوص الفائقة والوسائط الفائقة. ويبدو نهج البحث، في الواقع، قادراً على التغلب على المشكلات الثلاث التي تكتنف التصفح، كنهج لاسترجاع النصوص الفائقة والوسائط الفائقة. وبينما تشجع الوثائق فائقة البنيان على التصفح، فإن البحث يمكن أن يتغلب على أي مشكلات استرجاع، ترتبط بالتصفح في الفضاء الفائق. ومن ثم يبدو التكامل بين نهجي البحث والتصفح، أفضل سبيل لاسترجاع النصوص الفائقة والوسائط الفائقة.

المراجـع

Balkin, Ruth. (1999). AltaVista's automatic translation program. *Database,* 22(2), 56-57.

Barnard, Kobus, et al. (2003). Matching words and pictures. *Journal of Machine Learning Research,* 3,1107-1135.

Birmingham, William, Dannenberg, Roger, and Pardo, Bryan. (2006). Query, by humming with the VocalSearch system. *Communications of the ACM,* 49(8), 49-52.

Birmingham, William, et al. (2002). The MusArt music retrieval system. *D-Lib·Magazine,* 8(2). Retrieved December 19, 2008, from www.dlib.org/dlib/ february02/birmingham/02birmingham.html

Boykin, Stanley, and Merlino, Andrew. (2000). Machine learning of event segmentation for news on demand. *Communications of the ACM,* 43(2), 35-41.

Byrd, Donald, and Crawford, Tim. (2002). Problems of music information retrieval in the real world. *Information Processing & Management,* 38(2), 249-272.

Cawkell, A. E. (1992). Selected aspects of image processing and management: Review and future prospects. *Journal of Information Science,* 18,179-192.

Chang, Shih-Fu, et al. (1997a). VideoQ: An automated content based video search system using visual cues. *Proceedings of ACM Multimedia, Seattle, WA,* 313-324.

Chang, Shih-Fu, et al. (1997b). Visual information retrieval from large distributed online repositories. *Communications of the ACM,* 40(12), 63-71.

Chen, Aitao, and Gey, Fredric C. (2004). Multilingual information retrieval using machine translation, relevance feedback and decompounding. *Information Retrieval,* 7(1/2), 149-182.

Chu, Heting. (2001). Research in image indexing and retrieval as reflected in the literature. *Journal of the American Society for Information Science and Technology,* 52(12), 1011-1018.

CNNIC (China Internet Network Information Center). (January 2009). Statistical report on internet development in China. [In Chinese.] Retrieved January 20, 2009, from www.cnnic.net.cn/uploadfiles/pdf/ 2009/l/13/92458.pdf

Cornejo, J. G. (1994). *Hypertext: Its use in the documentary treatment of data.* Santiago, Chile: REDUCE-CIDE.

Coyle, Karen. (2005). Unicode: The universal character set. *The Journal of Academic Librarianship,* 31(6), 590-592.

Downie, J. Stephen. (2000). Access to music information: The state of the art. *Bulletin of the American Society for Information Science,* 26(5), 23-25.

Downie, J. Stephen. (2003). Music information retrieval. *Annual Review of Information Science and Technology,* 37,295-340.

Eakins, John P., and Graham, Margaret E. (January 1999). *Content-based image retrieval: A report to the JISC Technology Applications Programme.* Retrieved December 17,2008, from www.jisc.ac.uk/uploaded_documents/ jtap-039.doc

Enser, Peter. (2008). The evolution of visual information retrieval. *Journal of Information Science,* 34(4), 531-546.

Fallows, Deborah. (2007). China's online population explosion: What it may mean for the internet globally ... and for U.S. users. Pew Internet & American Life Project. Retrieved October 9,2009, from www.pewinternet. org/~/media/Files/Reports/2007/China_ Internet_July_ 2007.pdf.pdf

Faloutsos, C., et al. (1994). Efficient and effective query by image content. *Journal of Intelligent Information Systems,* 3,231-262.

Flickner, Maron, et al. (1995). Query by image and video content: The QBIC system. *IEEE Computer,* 28(9), 23-32.

Gauvain, Jean-Luc, Lamel, Lori, and Adda, Gilles. (2000). Transcribing broadcast news for audio and video indexing. *Communications of the ACM,* 43(2), 64-70.

Goodrum, Abby. (2000). Image information retrieval: An overview of current research. *Informing Science,* 3(2). Retrieved January 21, 2009, from inform.nu/Articles/Vol3/v3n2p63-66.pdf

Goodrum, Abby, and Rasmussen, Edie. (2000). Sound and speech in information retrieval: An introduction. *Bulletin of the American Society for Information Science,* 26(5), 16-18.

Gordon, Raymond G., Jr. (Ed.). (2005). *Ethnologue: Languages of the world.* 15th ed. Dallas, TX: SIL International.

Graham, Margaret E. (May 1999). *The description and indexing of images: Report of a survey of ARLIS members, 1998/99.* Retrieved February 1,2000, from www.unn.ac.uk/iidr/ARLIS

Gey, Fredric C., Kando, Noriko, and Peters, Carol. (2005). Cross-language information retrieval: The way ahead. *Information Processing & Management,* 41(3), 415-431.

Hauptmann, Alexander G., and Witbrock, Michael J. (1997). Informedia: News-on-demand multimedia information acquisition and retrieval. In Mark T. Maybury (Ed.), *Intelligent multimedia information retrieval* (pp. 213-239). Menlo Park, CA: AAAI Press.

Hollink, Vera, et al. (2004). Monolingual document retrieval for European languages. *Information Retrieval,* 7(1/2), 33-52.

Ingwersen, Peter. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation,* 52(1), 3-50.

Jorgensen, Corinne. (2003). *Image retrieval: Theory and research.* Lanham, MD: Scarecrow Press.

Kishida, Kazuaki. (2005). Technical issues of cross-language information retrieval: A review. *Information Processing & Management,* 41(3), 433-455.

Kubala, Francis, et al. (2000). Integrated technologies for indexing spoken languages. *Communications of the ACM,* 43(2), 48-56.

Lawrence, Steve, and Giles, C. Lee. (July 8,1999). Accessibility of information on the web. *Nature,* 400,107-109.

Lerner, Reuven M. (2003). At the forge: Unicode. *Linux Journal,* 2003(107). [Article No. 8, html file].

Lesk, Michael. (2005). *Understanding digital libraries.* 2nd ed. San Francisco: Morgan Kaufmann.

Lippincott, Aura. (2002). Issues in content-based music information retrieval. *Journal of Information Science,* 28(2), 137-142.

Lyman, Peter, et al. (2000). How much information? Berkeley, CA: School of Information Management and Systems, University of California at Berkeley. Retrieved January 21, 2009, from www.sims.berkeley.edu/how-much-info/summary.html

Miniwatts Marketing Group. (2009). Internet world users by language. Retrieved October 8,2009, from www.internetworldstats.com/stats7.htm

Mostafa, Javed. (1994). Digital image representation and access. *Annual Review of Information Science and Technology,* 29, 91-135.

Oard, Douglas W. (2000). User interface design for speech-based retrieval. *Bulletin of the American Society for Information Science,* 26(5), 20-22.

Panofsky, Erwin. (1955). *Meaning in the visual arts: Papers in and on art history.* New York: Doubleday Anchor Books.

Persson, Olle. (2000). Image indexing-A first author co-citation map. Retrieved January 21, 2009, from www.umu.se/inforsk/Imageindexing/ imageindex.htm

Rasmussen, Edie M. (1997). Indexing images. *Annual Review of Information Science and Technology,* 32,169-196.

Sands, Stella (Ed.). (2000). Languages. *Kids Discover,* 10(3), 3.

Schwartz, Barry. (2008). Search pictures with pictures at Live Search. Search Engine Land. Retrieved December 24,2008, from searchengineland.com/ search-pictures-with-pictures-at-live-search-15666.php

Si, Lou, et al. (2008). An effective and efficient results merging strategy for multilingual information retrieval. *Information Retrieval,* 11(1), 1-24.

Smeaton, Alan F. (2004). Indexing, browsing, and searching of digital video. *Annual Review of Information Science and Technology,* 38, 371-407.

Spärck Jones, Karen. (2000). Further reflections on TREC. *Information Processing & Management,* 36(1), 37-85.

Spärck Jones, Karen, et al. (1996). Experiments in spoken document retrieval. *Information Processing & Management,* 32, 399-419.

SRI MAESTRO Team. (2000). MAESTRO: Conductor of multimedia analysis technology. *Communications of the ACM,* 43(2), 57-63.

Stix, Gary. (2006). A farewell to keywords. *Scientific American,* 295(1), 91-93.

Trant, Jennifer. (2004). Image retrieval benchmark database service: A needs assessment and preliminary development plan. A Report Prepared for the Council on Library and Information Resources and the Coalition for Networked Information. Retrieved December 18, 2008, from www.clir.org/pubs/reports/trant04/tranttext.htm

Unicode, Inc. (2008). What is Unicode? Retrieved December 13, 2009, from www.unicode.org/standard/WhatIsUnicode.html

Voorhees, Ellen, and Garofolo, John. (2000). The TREC Spoken Document Retrieval track. *Bulletin of the American Society for Information Science,* 26(5), 18-19.

Voorhees, Ellen, and Harman, Donna. (2000). Overview of the Sixth Text REtrieval Conference (TREC-6). *Information Processing & Management,* 36(1), 3-35.

Wactlar, Howard D., et al. (2000). Complementary video and audio analysis for broadcast news archives. *Communications of the ACM,* 43(2), 42-47.

Wan, Gary, and Liu, Zao. (2008). Content-based information retrieval and digital libraries. *Information Technology & Libraries,* 27(1), 41-47.

الفصل العاشر

المستفيد طرفا في تنظيم المعلومات واسترجاعها

المستفيد عامل حاسم، ينبغي وضعه في الحسبان، في جميع أنشطة وجهود تنظيم المعلومات واسترجاعها؛ نظراً لأن الهدف النهائي لتنظيم المعلومات واسترجاعها هو تلبية احتياجات المستفيد من المعلومات، فالمعلومات، يتم تنظيمها على سبيل المثال بحيث يستطيع المستفيدون استرجاعها عند الحاجة إليها. وقد تم تطوير مختلف طرق الاسترجاع وأساليبه، بحيث يستطيع المستفيدون اختيار الملائم منها، بناء على احتياجاتهم المحددة من المعلومات.

ويعد المستفيد، أو ما يتصل بالمستفيد، في تنظيم المعلومات واسترجاعها، حيوي أياً كان المنظور الذي يمكن للمرء اتخاذه؛ فمن وجهة نظر مصمم النظام، على سبيل المثال، فإنه من أجل تلبية الاحتياجات المعلوماتية للمستفيدين، يتم بناء نظم تهيئة المعلومات واسترجاعها. وعلى الرغم من أن المتعاملين في تنظيم المعلومات اختصاصيو معلومات على الأرجح، يختلفون عن المستفيدين النهائيين في استرجاع المعلومات، فإنهم يتقاسمون معهم كثيرا من أوجه التشابه كمستفيدين في أنشطة تنظيم المعلومات واسترجاعها. ويعالج هذا الفصل المستفيدين واحتياجاتهم المعلوماتية، ثم يستكشف بعد ذلك النهج المعرفي الموجه نحو المستفيد، في تنظيم المعلومات واسترجاعها. وأخيرا نناقش التفاعل بين المستفيد ونظام استرجاع المعلومات، إلى جانب بعض القضايا المرتبطة بذلك.

المستفيدون واحتياجاتهم المعلوماتية:

المستفيدون أفراد، لكل منهم خصائصه المميزة. إلا أنه لا يمكن ولاشك دراسة كل مستفيد بمفرده، لأغراض تنظيم المعلومات واسترجاعها؛ فالواقع أن الباحثين يقسمون المستفيدين إلى فئات، بتطبيق معايير مشتركة، كالنوع، والسن، والمهنة، والظروف الاقتصادية، والثقافة، والشخصية، والخلفية التعليمية. وبناء على هذه الخصائص العامة يمكننا التعرف على احتياجات مختلف الفئات على نحو أفضل، حتى وإن كنا ندرك أن احتياجات المستفيد تتأثر بكثير من العوامل الأخرى أيضا (كمصادر المعلومات على سبيل المثال).

وإذا نظرنا إلى السن كمثال، فإن المستفيدين مختلفي الأعمار، لهم احتياجاتهم المعلوماتية المختلفة. فالمستفيدون في سن المدرسة، أو في ظروف العمل، بوجه عام، عادة ما تكون لديهم احتياجاتهم المعلوماتية بالنسبة للتعليم، أو المتصلة بالعمل، بينما تميل الفئات العمرية الأخرى للحاجة إلى المعلومات العامة في معظم الأحيان. والمهنة معيار آخر للتحقق من احتياجات المستفيد؛ فالمشتغلون بالعلم على سبيل المثال، يشكلون عادة ما يسمى الجامعات الافتراضية **invisible colleges** لتبادل المعلومات (راجع **Crane, 1972; Price, 1963** على سبيل المثال) بينما يعتمد المهندسون، في مختبرات البحث والتطوير، على السدنة **gatekeepers** للتواصل مع العالم الخارجي (راجع **Allen, 1970** على سبيل المثال). وقد استقصى شودري **Chowdhury (1999)** الاحتياجات المعلوماتية للمستفيدين، في مجالات أنشطة مختلفة (كإدارة الأعمال والمشروعات على سبيل المثال). وقد تم تلخيص المزيد من الدراسات حول سلوك المستفيدين في التماس المعلومات **information seeking behavior**، وهذا مصطلح أحدث، تبناه الباحثون في المجال ليحل محل التعبير الأقدم وهو الاحتياجات المعلوماتية وأوجه الإفادة، وذلك من جانب كيس **Case (2006, 2007)**، في دراسته

الوصفية التحليلية الضافية لمن يلتمسون المعلومات (أي المستفيدون)، وفقا للمهنة (كالمشتغلون بالعلم أو المهندسون على سبيل المثال)، والدور (كالمرضى والطلبة على سبيل المثال)، والجوانب الديموجرافية (كالسن والفئة الإثنية أو العرقية على سبيل المثال).

ولقد أعاد العصر الرقمي، ونمو الإنترنت على وجه الخصوص، النظر في تعريف المستفيدين واحتياجاتهم المعلوماتية، في عملية تنظيم المعلومات واسترجاعها فعلا. وهناك تزايد في جهود استكشاف خبرات البحث **search histories**، وأفضليات البحث (كنوع الملف، واللغة، ومجال الشبكة، والدولة)، وسياقات البحث الأخرى، وذلك في البحوث المتصلة، كاسترجاع المعلومات على الإنترنت (راجع على سبيل المثال **Komlodi, Soergel & Marchionini, 2006; Rose, 2006**). وفي الوقت نفسه، ينبغي أن يظل المستفيد متحكما في حدود السياق الذي يتضمنه استفسار البحث، لكل من دواعي الخصوصية، ولأن الخوارزميات الراهنة، قد لا تفسر السياق على نحو سليم في بعض الأحيان **(Resnick & Vaughan, 2006)**. ولما كان العصر الرقمي قد حل فعلا بكل تأكيد، وأنه قد وجد ليبقى؛ فإن احتمالات تأثير ذلك في أجيال المستقبل، في سبيله لأن يصبح موضوعا للبحث. وقد درس رولاندز وآخرون **Rowlands et al. (2008)**، على سبيل المثال، تأثير التحول الرقمي على السلوك المعلوماتي، لجيل جوجل (أي أولئك الذين ولدوا بعد عام ١٩٩٣). وتدعي الدراسة أنه على الرغم من أن الشباب يبدون سهولة واضحة، وتآلفا مع الحاسبات، فإنهم يعتمدون بكثافة على محركات البحث، ويشاهدون أكثر مما يقرأون، ويفتقرون إلى المهارات النقدية والتحليلية اللازمة لتقييم ما يجدون من معلومات على العنكبوتية.

وييسر الإلمام بالمستفيدين واحتياجاتهم، أنشطة تنظيم المعلومات واسترجاعها. وقد أجريت أعداد ضخمة من الدراسات، لاستكشاف تأثير سمات المستفيدين على

تنظيم المعلومات واسترجاعها في العصر الرقمي (مثل **Borgman, 1989; Buente & Robin, 2008; Fenichel, 1981; Marchionini, et al. 1993; Zhang & Chignell, 2001** بالنسبة لمن يقومون بإجراء عمليات البحـث، و **Faroog, et al., 2007; Kipp & Campbell, 2006** بالنـسبة للمكـشفين). ولقـد ازدادت درايتنـا بالمـستفيد في إطـار تنظـيم المعلومـات واسـترجاعها، ولاشـك بفـضل البحـوث من هذا النوع.

إلا أن الحاجة إلى المعلومات مفهوم غامض؛ إذ يفتقر إلى البنيان المنضبط، فضلا عن افتقاده للدقة وغياب التنظيم؛ ولهذا السبب أسماه بلكن وزملاؤه آسك **ASK**، أي الحالة المعرفية غير الـسوية **anomalous state of knowledge (Belkin, Oddy & Brooks, 1982)**. وفضلا عـن ذلـك فـإن للاحتياجـات المعلوماتيـة خـصائص أخرى، تشمل التغير بمرور الوقت، والاختلاف من مستفيد إلى آخر، والتوقف عـلى البيئة. يضاف إلى ذلك أن الاحتياجات المعلوماتية، غالبا ما تظل دون إعراب عنها، أو يكـون الإعـراب عنهـا بـشكل غـير مناسـب **(Chowdhury, 1999)**. والواقـع أن المستفيدين غالبا ما يجـدون صـعوبة في التعبـير عـن احتياجـاتهم المعلوماتيـة كاملـة، وناهيك عن التعبير عنها بدقة. وليس من غير المألوف أن تصدر العبارات التالية عـن المستفيدين، عندما يُسألون عما يبحثون عنه: "سوف أعرف عندما أراها". وغالبـا مـا يجري اختصاصيو المعلومات مقابلات مرجعية، وهي إحدى خطوات البحث الـذي يحظى بالمساعدة، للعمل على حل هذه المشكلة.

وفضلا عن خصائص المستفيد، تحقق بيزلي **Paisley (1968)** من أربعة عوامل أخرى تؤثر في الاحتياجات المعلوماتية للمستفيدين:

١. مصادر المعلومات: لمجموعة مصادر المعلومات المتاحة للمستفيدين بعض التأثير على احتياجاتهم. وتختلف توقعات المستفيدين عندما تتوافر لهم على

سبيل المثال، معلومات من نظام محلي، في مقابل ما يتوافر لهم من نظام وطني في تغطيته.

٢. الهدف: لأي هدف يحتاج المستفيدون إلى المعلومات؟ فإذا كان المستفيدون يرغبون في المعلومات لمجرد الإجابة عن بعض الأسئلة الحقائقية (ما أعلى مبنى في العالم؟ على سبيل المثال)، فإن احتياجاتهم المعلوماتية تبدو بسيطة أو مباشرة. أما إذا كان المستفيدون يريدون المعلومات من أجل مشروع بحث ما، فإن احتياجاتهم المعلوماتية تدخل ولاشك في فئة أخرى.

٣. العوامل الخارجية: يمكن للعوامل الاجتماعية، والسياسية، والاقتصادية، وغيرها من العوامل الخارجية، أن تؤثر بقوة في المستفيدين واحتياجاتهم المعلوماتية.

٤. النتيجة: ما الحصيلة التي تعود على المستفيدين، نتيجة للإفادة من المعلومات؟ فمن الممكن للنتائج الإيجابية، على سبيل المثال (زيادة الإنتاجية مثلا)، أن تكون بمثابة مكافآت للمستفيدين، كما تحثهم على المزيد في ذلك الصدد.

وقصارى القول، فإن المستفيدين أنفسهم، إلى جانب مصادر المعلومات، والهدف، والعوامل الخارجية، والنتيجة، يشكلون العوامل الخمسة التي تؤثر في الاحتياجات المعلوماتية. وعلى الرغم من أن بحث بيزلي **Paisley** حول الاحتياجات المعلوماتية، نشر عام ١٩٦٨، فإنه لا يزال صالحا اليوم، ونحن بصدد مناقشة الموضوع في سياق تنظيم المعلومات واسترجاعها رقميا.

ولما كانت الاحتياجات المعلوماتية تتأثر بكثير جدا من العوامل، فإنه من المتوقع أن تكون هناك أنواع مختلفة من هذه الاحتياجات. ومن التصنيفات المألوفة، تقسيم الاحتياجات المعلوماتية إلى فئتين؛ أولاهما الحاجة إلى عنصر معروف، والثانية

الاحتياجات الموضوعية (Lancaster & Warner, 1993)، وذلك بمجموعة أخرى من المصطلحات؛ كالاحتياجات المعلوماتية المحددة واضحة المعالم **concrete**، والاحتياجات المعلوماتية المرتبطة بمشكلة ما (Frants & Brush, 1988). وتعبر الحاجة إلى عنصر معروف، والاحتياجات المعلوماتية المحددة، عن مفهومين متشابهين، كما يصدق الأمر نفسه بالنسبة للاحتياجات الموضوعية، والاحتياجات المعلوماتية المرتبطة بمشكلة ما. فالفئة الأولى محددة واضحة المعالم، بينما الثانية ليست كذلك. وتتطلب تلبية الاحتياجات الموضوعية أو الاحتياجات المعلوماتية المرتبطة بمشكلة ما،بوجه عام، قدرا كبيرا من الجهد. ونادرا ما تطبق المعايير الأخرى (كفورية المعلومات على سبيل المثال) لتصنيف الاحتياجات المعلوماتية؛ فالمستفيدون في العلوم البحت **hard sciences**، على سبيل المثال، عادة ما يحتاجون إلى المعلومات الحديثة، بينما يمكن للمستفيدين في العلوم الهشة **soft** أن يكونوا أكثر اهتماما بالمعلومات الراجعة الشاملة.

وكما سبق أن بينا، فإن أنشطة تنظيم المعلومات واسترجاعها، ينبغي أن تمليها في المقام الأول الاحتياجات المعلوماتية للمستفيدين. ونظرا لأن المستفيد هو من يشعر بالحاجة إلى المعلومات، فإن بعد المستفيد، أو ما يتصل بالمستفيد في تنظيم المعلومات واسترجاعها، أمر واضح لا لبس فيه. ولهذا السبب يحاول الباحثون اتباع نهج يركز على المستفيد، لرسم مسار التطور في استرجاع المعلومات.

الأنموذج المعرفي والنماذج التي تركز على المستفيد:

عندما يكون التركيز في تنظيم المعلومات واسترجاعها، منصبا على المستفيد، يستعير الباحثون الذين يطورون النماذج والطرق الجديدة، نظريات ومناهج مستعارة من المجالات التخصصية المتصلة بالبشر، كعلم النفس، وعلم الاجتماع،

والأنثروبولوجيا أو علم الإنسان. فالأنموذج المعرفي لاسترجاع المعلومات، على سبيل المثال، إطار يستمد مقوماته بكثافة، من علم النفس المعرفي، بينما يستند تنظيم المعلومات واسترجاعها، في السياق، إلى أسس علم الاجتماع، والأنثروبولوجيا، والاتصال (**Rogers, 2004**). وتتناول الأقسام التالية الأنموذج المعرفي والنماذج الأخرى التي تركز على المستفيد في المجال.

الأنموذج المعرفي:

الإدراك **cognition** هو الملكة أو العملية الفكرية التي يتم بها الحصول على المعرفة **knowledge** أو اكتسابها، وذلك عن طريق الملاحظة، والتفكير، والحدس (**Soukhanov et al., 1984**). ويركز الأنموذج المعرفي، الذي يختلف عن النماذج الأخرى لاسترجاع المعلومات، على الأنشطة المعرفية للمستفيدين، أثناء عملية استرجاع المعلومات. كيف تتشكل، على سبيل المثال، الاحتياجات المعلوماتية للمستفيد؟ كيف يحكم المستفيد على صلاحية النتائج المسترجعة؟ وهذه الأسئلة صعبة ولاشك، ويعمل الباحثون على الإجابة عنها على مر السنين.

وتركز نماذج استرجاع المعلومات، التي تناولناها في الفصول السابقة، على المعلومات وجوانب نظام استرجاع المعلومات. فكيف يمكن، على سبيل المثال، التعبير عن المعلومات أو الاستفسارات؟ أي أساليب استرجاع المعلومات، ينبغي أن يحظى بالدعم؟ أي أساليب استرجاع المعلومات تعمل أفضل من غيرها؟ إلا أن هذه النماذج لا تزال محدودة على نحو ملحوظ؛ إذ إنها لا تولي السياقات الاجتماعية والمعرفية التي تتم فيها مهام التنظيم والاسترجاع، الاهتمام الكافي (**Ingwersen, 1996**). وبناء على أنموذج إنجورسن (**1996**) **Ingwersen** المعرفي للتفاعل في استرجاع المعلومات، هناك ثلاثة عناصر رئيسة (أهداف المعلومات، وتصميم نظام استرجاع المعلومات، والبيئة الاجتماعية أو المؤسساتية)، بالإضافة إلى الفضاء المعرفي

للمستفيد، الذي يشمل الحالة المعرفية الراهنة، والمشكلة أو الهدف، والشك، والحاجة المعلوماتية، والسلوك المعلوماتي. ويتم التركيز في الأنموذج المعرفي على العملية المعرفية للمستفيد.

وللأنموذج المعرفي الكثير من التطبيقات المختلفة؛ فقد صمم بلكن وأودي وبروكس **Belkin, Oddy and Brooks (1982)**، على سبيل المثال، نظاما تفاعليا لاسترجاع المعلومات (السؤال والإجابة على سبيل المثال) للكشف عن الاحتياجات المعلوماتية للمستفيد. وتحليل بروتوكولات التفكير بصوت مرتفع، أسلوب آخر، يطبق في فهم الأنشطة المعرفية للمستفيدين، عندما يكونون بصدد البحث عن المعلومات (راجع على سبيل المثال **Gorman, et al., 2002; Ingwersen, 1982**). وقد تناول بتيجرو وفايدل وبروس **Pettigrew, Fidel and Bruce (2001)**، الطرق المعرفية للبحث في سلوكيات البحث عن المعلومات، بالشرح المفصل، على الرغم من أن تعريفهم للأطر المعرفية، يبدو أوسع بكثير من التعريف المتبع في هذا الكتاب. وتكمن أكبر مظاهر قوة الأنموذج المعرفي، في تركيزه على إدراك المستفيد، ذلك الإدراك الذي يعد عاملا حاسما بالنسبة لعملية استرجاع المعلومات. كيف تتشكل الحاجة المعلوماتية؟ كيف تتحول إلى استفسار بحث؟ كيف يتم تقييم النتائج المسترجعة؟ وتتعلق كل هذه الأسئلة بإدراك المستفيد، ويهدف الأنموذج المعرفي لتقديم إجابات لمثل هذه الأسئلة.

والأنموذج المعرفي موجه نحو المستفيد، الأمر الذي يميزه عن النماذج الأخرى لاسترجاع المعلومات، التي تركز على مكوني أو عنصري المعلومات أو النظام، في عملية استرجاع المعلومات. فالمستفيدون يتسمون بالديناميكية، وكذلك الحال أيضا بالنسبة لاحتياجاتهم المعلوماتية؛ إذ إن لمختلف المستفيدين احتياجات معلوماتية مختلفة. ومن شأن الاحتياجات المعلوماتية للمستفيد نفسه، أن تختلف من وقت لآخر،

ومن مكان إلى آخر، بل إنها يمكن أن تتغير أثناء التفاعل مع النظام.كما أن الأنموذج المعرفي هو وحده القادر على التعامل مع ديناميكيات المستفيدين، فيما يتعلق بأنشطتهم المعرفية في استرجاع المعلومات. وينبغي أن يركز استرجاع المعلومات على المستفيد، ومن ثم، فإن التوجه نحو المستفيد في الأنموذج المعرفي، يمكن أن يسهم في بناء نظم مرنة، لاسترجاع المعلومات تراعي ظروف المستفيد.

وعلى الرغم من أن الأنموذج المعرفي يبدو واعدا إلى حد بعيد، بالنسبة للارتقاء بمستوى أداء استرجاع المعلومات، فإن فهم البشر للأنشطة المعرفية، لا يزال محدودا، فكما نبه إنجورسن **Ingwersen (1996)**:

> فإن الحاسبات (أو الكتب بالنسبة لهذا الأمر) تلتزم بمسلمات ثابتة محددة سلفا، بينما لا يمكن لأولئك البشر أن يتم التكهن بهم على المستوى الفردي؛ إذ أنهم يتشكلون كما هم بناء على الخبرات العارضة، والدلالية، والعاطفية. **(p. 6)**

والأنشطة المعرفية للمستفيدين في استرجاع المعلومات، أمر لا يمكن التكهن به؛ نظرًا لطبيعة الإدراك، ولأننا لم نستوعب بعد العملية المعرفية كاملة. وقد اقترح إنجورسن **Ingwersen (1996)** التنظيم المتعدد **poly-representation** للفضاء المعرفي للمستفيد، والفضاء المعلوماتي لنظم استرجاع المعلومات (كتوفير نقاط الوصول المتداخلة معرفيا، كالكلمات المفتاحية، والواصفات، والاستشهادات المرجعية، لأغراض الاسترجاع على سبيل المثال) كطريقة للتغلب على مشكلة الشك، وعدم القابلية للتنبؤ فيما يتعلق بالإدراك. وقد تكرر الاقتراح نفسه، من جانب كل من إنجورسن ويارفلين **Ingwersen and Jarvelin (2005)**.

وفضلا عن ذلك، فإن الإدراك ينطوي على بعض القيود الداخلية؛ فطاقة الذاكرة البشرية قصيرة المدى، على سبيل المثال، وهي سبع كتل **chunks** زائد أو

ناقص اثنتين (Miller, 1956). والكتلة مقياس لأصغر وحدة معلوماتية ذات دلالة. فرقم الهاتف في دولة ما، على سبيل المثال، يتكون من ثلاث كتل من المعلومات، هي رمز المنطقة (٢١٢ على سبيل المثال) ورمز الإقليم (٣٣٣ على سبيل المثال) والأرقام الأربعة الأخيرة (١٨١٩ على سبيل المثال). والمضمون المباشر لما انتهى إليه ملر **Miller**، بالنسبة للمستفيدين، في عملية استرجاع المعلومات، هو قدرتهم المحدودة على تجهيز المعلومات. فإذا كان أحد نظم استرجاع المعلومات يقدم اثني عشر خيارا مختلفا في الوقت نفسه، فإنه يمكن للمستفيدين مواجهة مشكلة؛ لأن مقدار المعلومات المقدمة يزيد عن طاقة استيعاب ذاكرتهم قصيرة المدى؛ ومن ثم فإن نظم استرجاع المعلومات، لا ينبغي أن يكون بها، على سبيل المثال، قائمة اختيار تشتمل على أكثر من تسعة خيارات. ومن ناحية أخرى يمكن للمستفيد أن يشعر أيضا بالإرهاق، أو الملل، أو القلق، عندما يقدم له الكثير من الاحتمالات. ويمكن لكل هذه العوامل أن تؤثر في الطاقة المعرفية أو الإدراكية للمستفيد. وينبغي أن يضع الأنموذج المعرفي لاسترجاع المعلومات، مثل هذه العوامل في الحسبان.

والبشر هم أكثر المخلوقات تعقدا في الكون. وأقل جوانب البشر قابلية للفهم هو إدراكهم، في حين يحاول الأنموذج المعرفي التصدي لمشكلة استرجاع المعلومات، من المنظور المعرفي. فالنهج مهم، وموجه فعلاً نحو الهدف مباشرة، إلا أنه يواجه أيضا تحديات ضخمة.

النماذج الأخرى لاسترجاع المعلومات التي تركز على المستفيد:

ازدهرت بحوث الاحتياجات المعلوماتية والمستفيدين، في السنوات الأخيرة. فوفقا لما انتهى إليه جوليان وداجان **Jullien and Duggan (2000)**، شكلت دراسات الاحتياجات المعلوماتية، والتماس المعلومات، والإفادة من المعلومات، ما يصل إلى حوالي ٨ بالمئة، من الإنتاج الفكري البحثي في علم المكتبات والمعلومات،

وقت إجراء دراستها. ومن المحتمل أن تكون هذه النسبة المئوية قد ارتفعت منذ ذلك الحين. وكثير من هذه الدراسات مكرس للنماذج والمداخل المفاهيمية، التي تركز على المستفيد في استرجاع المعلومات (راجع على سبيل المثال **Bates, 1989; Belkin et al., 1995; Dervin, 1992; Kuhlthau, 1993; Wilson, 1999**). ويمكن الاطلاع على مراجعات إضافية لهذه الدراسات في كيس **Case (2007)**، وانجورسن ويرفلين **Ingwersen and Järvelin (2005**، وسبنك وكول **Spink and Cole (2006)**، وولصون **Wilson (1999)**، إذ استقصى المؤلفون، وقارنوا واستكشفوا نماذج متعددة، تركز على المستفيد، وعلى نحو لا يستهان به من التفصيل.

ولا تتسم تلك النماذج التي تركز على المستفيد بالوفرة في الكم فحسب، وإنما تشكلت أيضا بتوجهات متنوعة أيضا، ومن زوايا نظر مختلفة. وبذلك يحاول الباحثون تقسيم مثل هذه النماذج إلى فئات، وفقا لمعايير معينة، من أجل مساعدة الآخرين على تحقيق فهم أفضل لها. فقد ذهب يرفلين وولصون **Järvelin and Wilson (2003)**، على سبيل المثال، إلى أن بعض النماذج ذات طابع موجز (راجع على سبيل المثال **Ellis, 1989; Ingwersen, 1996**) بينما النماذج الأخرى أكثر تحليلا (راجع على سبيل المثال **Belkin, Oddy, & Brooks, 1982; Bystron & Järvelin, 1995**). ومن ناحية أخرى قسم سبنك وكول **Spink and Cole (2006)** النماذج التي تركز على المستفيد، إلى ثلاث فئات؛ هي منظور حل المشكلات (راجع على سبيل المثال **Belkin, Oddy & Brooks, 1982; Kuhlthau, 1993; Wilson, 1999**)، والتماس معلومات الحياة اليومية، وإضفاء المعنى (راجع على سبيل المثال **Dervin, 1992; Huotari & Chatman, 2001**)، والتزود بالمعلومات (راجع على سبيل ا لمثال **Bates, 1989; Pirolli & Card, 1999**). وبصرف النظر عن الطريقة التي يتم بها اختبار النماذج فيما بعد، فإنها جميعها تتسم بخاصية واحدة مشتركة؛ ألا وهي وضع المستفيد في مركز الأنموذج الذي يتم تصميمه.

وقد أشار ولـصون (Wilson (2008 في وقـت لاحـق، إلى أن تطـور بحـوث استرجاع المعلومات، التي تركز على المستفيد، قـد أفـضى إلى التمييـز بـين احتياجـات الأوساط الأكاديمية إلى أعمال تـستند إلى أسـاس نظـري، واحتياجـات الممارسـين إلى توجيهات الارتقاء بمستوى الخدمات. كذلك بيَّن ولـصون أن الفـصل بـين البحـث والممارسة، يتحقق إلى مدى لا يستهان به؛ فقد كانت البحوث المبكرة يجريها ممارسـون، أما البحوث الراهنة فيتم إجراؤها فى الوقت الذي تسيطر فيه الأوساط الأكاديمية على المشهد. إلا أن الفصل بين البحث والممارسة يمكـن أن يـؤدي إلى مخـاوف في المجـال؛ نظرا لأن الأنموذج الذي يركز على المستفيد، يمكن أن يكون بلا قيمة، مـالم تكـن لـه مضامين تطبيقية يعتد بها في استرجاع المعلومات، مهما كان من الممكن أن تتوافر له من مقومات الروعة. ويميل الفصل بين البحث والتطبيق أو الممارسة، لأن يفضي إلى نتائج لا أهمية لها، في مجال استرجاع المعلومات. يضاف إلى ذلك أنه ربما يكون قـد آن الأوان لتأمل النماذج التي تركز على المستفيد، التي صممت فعلا، والبحث في السبل المحتملة لتطبيقها، لمـساعدة المـستفيدين عـلى الارتفـاع بمـستوى أدائهـم، في أنـشطة الـتماس المعلومات واسترجاعها. وبينما يبذل بعض الباحثين (راجع على سـبيل المثـال ,Case (2007; Ingwersen & Järvelin, 2005; Wilson, 1999 جهـودا طيبـة في هـذا الصدد، تدعو الحاجة، ولاشك إلى المزيد من العمل في هذا المجال سريع التطور.

تفاعل المستفيد مع النظام:

واجهة المستفيد user interface، مصطلح تم سكه لوضع ما يتعلق بالمستفيد على وجه التحديد، في تنظيم المعلومات واسترجاعها في الحسبان؛ فواجهة التعامل هي ما يراه المستفيد ويـسمعه ويلمـسه، في التفاعـل مـع النظـام (Shaw, 1991). إلا أن مصطلح الواجهة يبدو ساكناً static، ولا يركز إلا على نقطة واحدة لتفاعل المـستفيد

مع النظام. ولهذا السبب نستعمل مصطلح "تفاعل المستفيد مع النظام" **user-system interaction**، أو "التفاعل" **interaction** ببساطة، في هذا الكتاب، للدلالة على الأنشطة الديناميكية التي تدور بين المستفيد ونظام استرجاع المعلومات. والمصطلح الآخر الذي ينبغي أن يحظى ببعض التوضيح، في هذا الفصل، هو تفاعل الإنسان مع الحاسب **human-computer interaction**، أو تفاعل الحاسب مع الإنسان **computer-human interaction**. ولم يستعمل أي من المصطلحين في هذا الكتاب؛ لأن كلا من "الحاسب" و"الإنسان" لهما معانٍ أوسع مما يتعلق بالنظام والمستفيد، في تنظيم المعلومات واسترجاعها، أي المجال الذي يهدف هذا الكتاب إلى تغطيته.

أساليب تفاعل المستفيد مع النظام:

يمكن لتفاعل المستفيد مع النظام، في تنظيم المعلومات واسترجاعها، أن يحدث بأي من الأساليب التالية: لغة إصدار الأوامر **command language**، والاختيار من قائمة الخيارات **menu selection**، والتشغيل التصويري **graphical operation**، وتعبئة النماذج أو الاستمارات **form filling**، والروابط الفائقة **hyperlinks**، وحوار اللغة الطبيعية **natural language dialog**. ومن الممكن أيضا تطبيق هذه الأساليب، تزامنيا، أي بالتوازي لتنظيم المعلومات واسترجاعها.

لغة إصدار الأوامر:

لقد كانت لغة إصدار الأوامر هي الشكل الأول زمنيا، الذي تم وضعه لتفاعل المستفيد مع النظام، وكانت لهذا الشكل السيطرة في نظم استرجاع المعلومات، وخاصة نظم الخط المباشر، مثل ديالوج **DIALOG**، في العهود المبكرة لتنظيم المعلومات واسترجاعها. وفي هذا الأسلوب يضغط المستفيد على مفاتيح لغة إصدار الأوامر (أي

اللغة الاصطناعية المصممة خصوصًا لأغراض الاسترجاع)؛ وذلك لإعطاء الأوامر للنظام لإنجاز وظائف معينة (كالبدء والبحث على سبيل المثال). ومصطلحات لغة إصدار الأوامر موجزة، مثل "تسجيل الخروج" **logoff** لإنهاء عملية البحث وترك نظام الاسترجاع. والميزة الأخرى للغة إصدار الأوامر، هي سرعة التفاعل **(Large, Tedd & Hartley, 1999)**. وبالمقارنة بأشكال التفاعل الأخرى، فإن لغة إصدار الأوامر أسرع بكثير، في الحصول على الاستجابات من النظام؛ لأنه ليست هناك أي خطوات إضافية (كترجمة قائمة الخيارات إلى أوامر للنظام على سبيل المثال) يتم تنفيذها. وبذلك المعنى يكون التفاعل بين المستفيد والنظام مباشرا. كما أن للسرعة قيمتها أيضا، في بيئة تتوقف فيها رسوم الاسترجاع إلى حد بعيد، على وقت الارتباط وتكلفة الاتصالات. وفضلا عن ذلك يجد المستفيدون مرونة في اختيار أي الأوامر يمكن استخدامها، ولا يقتصرون على الخيارات التي ترد في قائمة الخيارات على سبيل المثال. ونتيجة لذلك يستطيع المستفيدون التحكم في التفاعل على نحو مباشر.

غير أن لغة إصدار الأوامر، كأسلوب للتفاعل، لا تبدي تعاطفا مع المستفيد، وعلى المستفيد أن يتعلم هذه اللغة، وأن يتمرس بها، ويطبقها على نحو منتظم، نظرا لطابعها الاصطناعي. ويمكن لمعنى الأمر ألا يكون في بعض الأحيان واضحا وصريحا بالنسبة للمستفيد؛ فكلمة "اختر" **select**، على سبيل المثال، تقوم مقام "ابحث عن" **search for**، في لغة ديالوج. يضاف إلى ذلك أنه قلما يكون هناك توحيد **uniformity** بين لغات إصدار الأوامر، من حيث النحو أو النظم **syntax** والدلالة. فمن الممكن للغة إصدار الأوامر، بالنسبة لإنهاء جلسة البحث، أن تشتمل على "خروج" **exit** في نظام ما، و"تسجيل الخروج" **logoff** في نظام آخر. ومهما يكن فإن التفاعل القائم على لغة إصدار الأوامر قلما يكفل للمستفيد المساعدة والإرشاد، فيما يتعلق بما يتعين عليه القيام به. ويمكن للنظام على سبيل المثال، أن يستعجل "أدخل

أمرًا enter a command" ولا يلمح بشأن أي أمر يمكن إدخاله. ونظرا لهذه الصعوبات، فإن لغة إصدار الأوامر غالبا ما تكون أسلوب التفاعل بين نظام استرجاع المعلومات والوسطاء (اختصاصيو المكتبات واختصاصيو إجراء عمليات البحث على سبيل المثال) لا بين نظام استرجاع المعلومات والمستفيدين النهائيين.

الاختيار من قائمة الخيارات:

الاختيار من قائمة الخيارات، أسلوب للتفاعل صمم جزئيا للتغلب على أوجه قصور لغة إصدار الأوامر. ولقد كانت نظم الأسطوانات المكتنزة، والفهارس المتاحة على الخط المباشر، أول أنواع نظم استرجاع المعلومات التي تستخدم أسلوب التفاعل المعتمد على الاختيار من قائمة الخيارات؛ إذ يتفاعل المستفيد مع النظام بالاختيار من الخيارات المتاحة، المسجلة في قائمة الخيارات، الأمر الذي يفيد من قدرات البشر بالنسبة للإدراك، التي تفوق قدرات التذكر. ومن ثم، فإن المستفيد في أسلوب الاختيار من قائمة الخيارات، لم يعد بحاجة لأن يتذكر الأوامر والعمل بها، كما في أسلوب لغة إصدار الأوامر. كذلك تتسم قوائم الخيارات بالتحديد بالنسبة لعملية الاسترجاع. وتختلف خيارات القائمة التي تظهر، على سبيل المثال، أثناء جلسة البحث، عن تلك التي تظهر وقت مشاهدة نتائج عملية البحث. وفي مقابل ذلك على المستفيد أن يقرر متى يستعمل أي أمر، في أسلوب لغة إصدار الأوامر. يضاف إلى ذلك أن المساعدة تظهر في شكل قائمة الخيارات في متناول المستفيد، في أسلوب التفاعل القائم على قائمة الخيارات.

وكأسلوب للتفاعل، يبدو الاختيار من قائمة الخيارات، أكثر تعاطفا أو أكثر مراعاة لظروف المستفيد، مما هو عليه الحال في لغة إصدار الأوامر. إلا أن الاختيار من قائمة الخيارات لا يتمتع بسرعة لغة إصدار الأوامر؛ إذ أنه يتعين على المستفيد أن يتفاعل مع النظام عن طريق قائمة الخيارات، وقد يستغرق الاختيار من قائمة

الخيارات وقتا أطول من كتابة الأمر، وخصوصا عندما يتوافر بالنظام عدة مستويات لقوائم الخيارات. ومن الممكن أن يقع العبء الزائد على الذاكرة، إذا كان عدد خيارات القائمة يزيد عن طاقة الذاكرة قصيرة المدى (أي رقم ملر **Miller** السحري سبعة زائد أو ناقص اثنين، الذي تناولناه في ١٠/ ٢/ ١) كذلك يفتقد المستفيدون المرونة في التفاعل مباشرة مع النظام؛ نظرا لأن اختيارهم يقتصر على ما يمكن أن يوفره النظام. إلا أنه بالنسبة للمستفيدين المبتدئين يفضل أسلوب الاختيار من قائمة الخيارات، عند التفاعل مع نظم استرجاع المعلومات.

الأسلوب التصويري للتفاعل:

يتم تطبيق الأسلوب التصويري لتفاعل المستفيد مع النظام، بواسطة الأيقونات **icons** أو الرموز التصويرية، والأزرار (بما في ذلك أزرار المذياع) والنوافذ، والخرائط التي يمكن النقر عليها. ويتفاعل المستفيد مع النظام بالتعامل (بالإشارة والنقر على سبيل المثال) مع مثل هذه التمثيلات البصرية. ويعتمد الأسلوب التصويري للتفاعل على الرموز في نظم استرجاع الإنترنت. ويستخدم أسلوب التفاعل هذا تمثيلات لموضوعات الكلمات الحقيقية (كأيقونة الطابعة للطباعة، وأيقونة الأسطوانة لحفظ الملفات، على سبيل المثال)، ووسيمات المفاهيم (كالمفتاح الموسوم للبحث على سبيل المثال) بوصفه الواجهة بين المستفيد والنظام. ومن الممكن للنوافذ أن تفتح، ويعدل حجمها، وتغلق، تبعا للحاجة بالنسبة للتفاعل. وتكفل الخرائط القابلة للنقر، للمستفيد استرجاع المعلومات، بمجرد النقر على النقطة التي تمثلها. فمن الممكن، على سبيل المثال، لنقرة على مدينة نيويورك، على خريطة وطنية للولايات المتحدة الأمريكية، أن تفضي إلى عرض خريطة المدينة. ويمكن لخاصتي التكبير والتصغير، إذ كانتا تحظيان بالدعم، أن تكونا مفيدتين إلى أبعد حد، في التعامل مع الخرائط القابلة للنقر.

ويمكن للأسلوب التصويري للتفاعل أن يتخذ أشكالا مختلفة، إلا أن هذا الأسلوب بوجه عام، يكفل للمستفيدين اليسر عندما يتفاعلون به؛ فتصوير الأنشطة التي تنطوي عليها مهام الاسترجاع، يؤدي إلى تجنب زيادة تحميل المستفيد بالأعباء المعرفية (كالاختيار من بين قائمة طويلة من الخيارات، على سبيل المثال). ويمكن للمستفيد أن يجد متعة في أسلوب التفاعل التصويري، الحي والملون في غالب الأحيان. وتكفل التعددية الكامنة في هذا الأسلوب، للمستفيد القدرة على إنجاز مهمة الاسترجاع على نحو أكثر كفاءة؛ فمن الممكن، على سبيل المثال، للمستفيد أن يختار أيقونات للتعامل مع المهام المحددة واضحة المعالم، في استرجاع المعلومات (كالطباعة على سبيل المثال) بينما يختار المفاتيح للأمور المجردة (كالبحث البوليني على سبيل المثال).

إلا أنه لا ينبغي الاستهانة بالصعوبة التي ينطوي عليها تطوير الأسلوب التصويري للتفاعل؛ فبعض العمليات في استرجاع المعلومات (كالاتجاه نحو قمة الوثيقة، على سبيل المثال) لا يمكن تمثيلها على النحو المناسب بالأيقونات. يضاف إلى ذلك أنه من الممكن للتفاعل التصويري أن يكون غير ضروري بالنسبة لأنشطة معينة في استرجاع المعلومات (كإدخال الاستفسارات على سبيل المثال). ويمكن لمعنى التمثيلات التصويرية، مالم يكن قد تم اختياره على نحو مناسب، أن يكون مضللا أو يساء فهمه. وتشغل التصميمات التصويرية للتفاعل المزيد من حيز الشاشة، ومن ثم فإنها قد لا تفسح مجالا للمعلومات القيمة **(Shneiderman & Plaisant, 2005)**. كما أن الوقت اللازم للتفاعل في الأسلوب التصويري، أطول بكثير مما هو عليه في أسلوب لغة إصدار الأوامر، أو أسلوب الاختيار من قائمة الخيارات، نظرا للمواد التصويرية. ومن ثم، فإنه يمكن لاسترجاع المعلومات بهذا الأسلوب أن يكون بطيئا وثقيلا، وهو أمر قد لا يثير مخاوف المستفيد؛ لأن وقت الارتباط لا يشكل سوى القليل في حساب تكلفة الاسترجاع في مثل هذه النظم لاسترجاع المعلومات (كنظم استرجاع الإنترنت على سبيل المثال).

النماذج الأخرى لتفاعل المستفيد مع النظام:

الاستمارات أحد أساليب التفاعل بين المستفيد والنظام؛ إذ يقوم المستفيد عند اتباع هذا الأسلوب بتسجيل عناصر كمصطلحات البحث، ويرسم حدود البحث، على استمارة يوفرها النظام. وينطوي أسلوب تعبئة الاستمارات فعلا على خواص قوائم الخيارات المنسدلة **pull-down** والأزرار. والفهارس المتاحة على الخط المباشر، ونظم استرجاع الإنترنت، نوعان من نظم استرجاع المعلومات التي غالبا ما تدعم أسلوب التفاعل هذا. فالمستفيد يقوم بإجراء عملية البحث بالاطلاع وتعبئة الاستمارة. وليست هناك قرارات صعبة يتخذها المستفيد، وإنما يرشد النظام المستفيد خلال عملية صياغة الاستفسار وتقديمه. ووجه القصور الواضح في هذا الأسلوب، هو افتقاده للمرونة في تفاعل المستفيد مع النظام؛ نظرا لأن المستفيد مقيد بما يرد في الاستمارة.

وتمثل الروابط الفائقة شكلا آخر للتفاعل. وهناك بعض أوجه التشابه بين الروابط الفائقة والاختيار من قائمة الخيارات؛ حيث يتفاعل المستفيدون بهذين الأسلوبين بالنقر على أحد عناصر قائمة الخيارات، أو إحدى الروابط الفائقة، التي تفضي إلى كيان آخر (قائمة خيارات أو إحدى الوثائق الأخرى على سبيل المثال). إلا أن الروابط الفائقة لا يتعين تقديمها في مجموعات، كما في حالة الاختيار من قائمة الخيارات. ومن الممكن للروابط الفائقة أن تنشأ حيثما تدعو الحاجة إليها، منفردة أو مع روابط فائقة أخرى. وكما بينا في ٦/ ٢/ ٢ فإن بنيان الرابطة الفائقة، يشجع على التصفح؛ ومن ثم فإن الروابط الفائقة تطبق على نطاق واسع، في نظم استرجاع المعلومات، التي تركز على التصفح. وياهو! **Yahoo!**، نظام استرجاع الإنترنت القائم على الأدلة، مثال ممتاز لأسلوب التفاعل القائم على الروابط الفائقة. ويقوم المستفيد بإنجاز مهام الاسترجاع بالتصفح عبر الفئات المرتبطة ارتباطا فائقا، التي يطورها

ياهو!. ويكمن وجه القصور الرئيس للروابط الفائقة كأسلوب للتفاعل، في بنيانها الشبكي. فمن الممكن للمستفيد أن يضل طريقه بسهولة عند التفاعل مع النظام باتباع الروابط؛ لأن الروابط الفائقة تتبع البنيان الشبكي، وغالبا ما يشبه البنيان الداخلي للروابط الفائقة المتاهة، لا المجموعة محكمة البنيان من الروابط. ومما لا شك فيه، أنه كلما زاد عدد الروابط الفائقة في نظام الاسترجاع، أصبح النظام أكثر إرباكا بالنسبة للمستفيد.

وقد بدأت اللغة الطبيعية تستخدم في التفاعل بين المستفيد والنظام، ويبدو تقديم الاستفسارات باللغة الطبيعية لنظام استرجاع المعلومات، أحد الجهود الضخمة في هذا الصدد. فنظام آسك دوت كوم **Ask.com** (الذي كان يسمى من قبل سل جيفز **Ask Jeeves**) يكفل للمستفيد إدخال الاستفسارات، باستعمال الجمل الكاملة (مثل: ما معالجة اللغة الطبيعية؟). وأكبر مزايا استعمال اللغة الطبيعية، في التفاعل بين المستفيد والنظام، هي طابعها الصريح المباشر؛ إذ يمكن لأي تمثيل للاستفسار أن يسفر حتما عن تحريف وتغيير في المعنى؛ ومن ثم فإن المستفيد في أسلوب اللغة الطبيعية للتفاعل، قد لا يكون بحاجة لأن يفسر معنى التمثيلات التي تنشأ لأغراض التفاعل (مثل لغة إصدار الأوامر، والأيقونات، وعناصر قائمة الخيارات). وبينما تبدو اللغة الطبيعية شكلا مرغوبا فيه إلى أقصى حد للتفاعل بين المستفيد والنظام، فإن التطبيق الناجح لها، يتوقف في النهاية على تقدم البحوث حول معالجة اللغة الطبيعية، وهذا موضوع نستكشفه في الفصل الثاني عشر.

ويرتبط ارتباطا وثيقا بأسلوب اللغة الطبيعية للتفاعل، استخدام الصوت في نظم استرجاع المعلومات؛ فالمستفيد في هذه الحالة يتفاعل مع النظام عن طريق اللغة المنطوقة لا اللغة المكتوبة. ويعتمد هذا الأسلوب للتفاعل بقوة، على تقنيات التعرف على الأصوات، التي لا تزال في مرحلة التطور.

وقد بدأت واجهة المستفيد العضوية **organic user interface (OUI)** تظهر في الإنتاج الفكري (راجع على سبيل المثال **Vertegaal & Poupyrev, 2008**) كشكل محتمل لتفاعل المستفيد مع النظام. وقد اختير اسم هذا الشكل لأن: (١) التقنيات الداعمة بالنسبة لبعض أهم تطورات واجهة المستفيد العضوية، هي الإلكترونيات العضوية، و(٢) أن ملايين الأشكال العضوية التي نلاحظها في الطبيعة، مذهلة في تنوعها، ومرونتها، وقابليتها للنقل أو التحول، وتكيفها الطبيعي، واحتمالاتها التطورية، بينما تبدو مرنة ومن الممكن الاعتماد عليها في الوقت نفسه. وقد أحصى فرتجال وبوبيريف **Vertegaal and Poupyrev (2008)** ثلاثة جوانب، تحدد معالم واجهة المستفيد العضوية؛ أولها أن المدخلات تساوي المخرجات، ووسيلة العرض أيضا هي وسيلة المدخلات. وشاشة اللمس بمثابة مثال في هذه الحالة. ويقدم ركيموتو **Rekimoto (2008)** المزيد من التوضيح لهذا الجانب لواجهة المستفيد العضوية، بأمثلة تفصيلية، مثل هولووول **Holowall**، وسمارتسكين **Smartskin**.

وثاني هذه الجوانب أن الوظيفة تساوي الشكل؛ إذ يمكن للعرض أن يتخذ أي شكل. وتحل واجهة المستفيد العضوية محل المصفوفة المستطيلة الجامدة، بالمزيد من المفاهيم المائعة **fluid**، التي تبدو أسطح رقيقة، مرنة ولمسية **tactile**، مهيأة للشكل والحيز **(Co & Pashinkov, 2008)**. ويمكن لجهاز قراءة الكتب الإلكترونية الذي يمكن طيه، أن يصبح حقيقة واقعة، باستخدام تقنيات عرض واجهة المستفيد العضوية الناشئة.

أما الجانب الثالث لواجهة المستفيد العضوية، فهو ذلك الشكل متواصل التدفق **follows flow** (أي تدفق تفاعل المستفيد): وبإمكان العروض تغيير شكلها لإنجاز الوظائف اللازمة، أثناء تفاعل المستفيد مع النظام. وما ينطوي عليه هذا الجانب هو ما يسمى التصميم الحركي، الذي يتم من خلاله إدخال القابلية للنقل الفيزيائي في الواجهة **(Parkes, Poupyrev & Ishii, 2008)**. ولما كانت واجهة المستفيد العضوية لا تزال رؤية لم تتحقق بعد، فإنه من السابق لأوانه كثيرا، معرفة ما لها وما عليها كأسلوب للتفاعل بين المستفيد والنظام.

الأسلوب الهجين للتفاعل:

شهدت السنوات القريبة الماضية ظهور أسلوب هجين للتفاعل، وخصوصا في نظم استرجاع الإنترنت. ومن الممكن لأي تجمع من أساليب التفاعل التي سبقت معالجتها، أن يشكل واجهة هجينا، على الرغم من أن لغة إصدار الأوامر قلما تندمج، نظرا لما يكتنفها من أوجه القصور؛ فإن نظام إكزاليد **Exalead**، وهو نظام لاسترجاع الإنترنت، مقره فرنسا، يدعم الاختيار من قوائم الخيارات، والروابط الفائقة، ولغة إصدار الأوامر (بالنسبة للمعاملات البولينية، ومعاملات التقارب وما شابه ذلك) فضلا عن الواجهة التصويرية. ويمكن للأسلوب الهجين للتفاعل، أن يقضي على أوجه القصور التي تكتنف أسلوبا بعينه، ويرتفع بمستوى كفاءته ومراعاته لظروف المستفيد. والواقع أننا قلما نشهد التفاعل أحادي الأسلوب، في نظم استرجاع المعلومات التي تطورت حديثا؛ نظرا لأن لكل أسلوب للتفاعل ما يكتنفه من أوجه القصور.

الأبعاد الأخرى لتفاعل المستفيد مع النظام:

فضلا عن أسلوب التفاعل، هناك جوانب أخرى ينبغي أن يضعها المستفيد في الحسبان، عند التفاعل مع نظام استرجاع المعلومات. وتشمل هذه الأبعاد خصائص العرض، وخيارات المخرجات، ومقومات المساعدة.

خصائص العرض:

يتفاعل المستفيد مع نظام استرجاع المعلومات عن طريق شكل معين من العرض، إلا أن الأدوات المادية اللازمة للتفاعل (كلوحة المفاتيح أو الفأرة على سبيل المثال) تخرج عن حدود مجال هذا الكتاب. ومن الممكن للتفاعل أن تتحدد معالمه بمثل خصائص العرض هذه، كاللون، وأشكال الحروف، والكثافة، وإخراج الشاشة.

ويوضح إخراج الشاشة كيف يتم عرض المعلومات اللازمة لأغراض التفاعل، على واجهة التعامل (شاشة حاسب عادة)؛ فهل يدعم الإخراج الأطر المتراكبة، أو النوافذ المتعددة؟ هل يحتاج المستفيد إلى التوجه إلى أعلى أو إلى أسفل ليرى كل المعلومات المقدمة؟ ومن الممكن للاستخدام المكثف لخاصية التراكب **overlaying**، أو الأعداد الكثيرة من الأطر أو النوافذ، في عرض واحد، أن يؤثر في قدرة المستفيد على التفاعل مع النظام بكفاءة. وكذلك الحال أيضا، فإن معظم المستفيدين (المبتدئين على وجه الخصوص) غير متمرسين في أساليب التنقل أو اللف **scrolling**. ومن المفضل دائما عرض كل المعلومات على شاشة واحدة، إن أمكن، لا على شاشات متعددة.

ويحكم نوع الخط وأسلوبه وحجمه، جودة عرض المعلومات النصية. والحد الأدنى للمواصفات بالنسبة لعرض النصوص، هو الوضوح والقابلية للقراءة. ويرى جاليتز **Galitz (1997)** أنه لا يمكن استعمال سوى نوعين من الخطوط (مثل الروماني الجديد للتايمز **Times New Roman**، أو الساعي **Courier**) وشكلين من الأحرف (مثل المألوف **regular** والمائل **italic**)، وثلاثة أحجام، للعرض الواحد. وينبغي تجنب أنواع الخطوط غير الواضحة أو غير المقروءة، أو أحجام الخطوط متناهية الصغر. ويمكن لنمط الخط إذا ما استعمل على النحو الملائم، أن ييسر عملية التفاعل.

ويتحكم اللون، والتباين، واللمعان، في بعد آخر للعرض، وهو على وجه التحديد المستوى المريح للمستفيد، عند التفاعل مع النظام. ويحدد عدد الألوان وتجميع أو خليط الألوان ما إذا كان العرض النهائي يبدو مستساغا للمستفيد أم لا. وينبغي أيضا اختبار الدرجات المناسبة للتباين واللمعان؛ لأن الإفراط في أي من العاملين، يمكن أن يتسبب في عدم راحة أعين المستفيدين. ويدخل لون الخلفية،

وصورة الخلفية في هذه الفئة من خصائص العرض. ويتبين من الخبرة أن الأسود على الأبيض، تجمع جيد بوجه عام بالنسبة للعرض. وينبغي مراعاة الحيطة والحذر عندما تستخدم الصور كخلفيات؛ نظرا لأنه من الممكن للصور أن تلقي بظلالها على المعلومات المهمة الأخرى التي ينبغي عرضها. كذلك يمكن لتحميل مثل هذا العرض أن يستغرق وقتا طويلا. ولهذه الأسباب، فإن بعض نظم استرجاع الإنترنت، التي كانت تستخدم خلفيات من الصور في أيامها المبكرة، سرعان ما أقلعت عن هذا الأسلوب.

وكثافة العرض عامل آخر، يمكن أن يكون له تأثيره على سهولة التفاعل بين المستفيد والنظام؛ فكم من المعلومات على سبيل المثال، يمكن أن يعرض على شاشة واحدة؟ وبينما تبدو المواصفات التفصيلية حول الكثافة غير محتملة؛ نظرا لأن عوامل كمجموعات الأحرف لها تأثيرها أيضا على الكثافة، فإنه ينبغي مراعاة الدرجة القصوى للكثافة في موقف معين. ولا يمكن بوجه عام لأي من الشاشة المزدحمة أو الفراغ الزائد عن الحد، أن يحظى بالترحيب من جانب المستفيد.

ومن الممكن للمؤثرات الخاصة، كتسليط الضوء و الوميض، والعرض المرئي العكسي **invese video**، أن تعمل جميعها على تحقيق نتائج معينة في تفاعل المستفيد مع النظام، إذا ما استخدمت على نحو سليم. ومن الممكن للاستخدام غير المطرد، أو الإسراف في استخدام أي مؤثر خاص، أن يؤدي إلى الارتباك وربما الإزعاج بالنسبة للمستفيد. ومن بين جميع الأنواع الرئيسة لنظم استرجاع المعلومات، تبدو نظم استرجاع الإنترنت الأكثر تعرضا للشاشات المعقدة؛ نظراً للإفراط في استعمال المؤثرات الخاصة، والخواص الأخرى للعرض.

خيارات المخرجات:

من الممكن تقديم نتائج الاسترجاع بأشكال مختلفة، بناء على محددات كالصيغ والكميات. وصيغ المخرجات في نظم الخط المباشر، هي الأكثر تنوعا؛ إذ يمكن للمستفيد أن يختار للمخرجات من بين صيغ محددة المعالم سلفا (كالمعلومات الوراقية فقط، أو المعلومات الوراقية بالإضافة إلى المستخلصات على سبيل المثال) أو يحدد حقولا بعينها (كالمؤلف، والعنوان على سبيل المثال). وعادة ما تدعم نظم الفهارس المتاحة على الخط المباشر ثلاث صيغ للمخرجات؛ الصيغ الموجزة، والطويلة، وتسجيلات مارك. أما فيما يتعلق بنظم استرجاع الإنترنت، فإن خياراتها لصيغ المخرجات محدودة؛ نظرا لطريقة تكشيفها. وعادة ما تتكون نتائج الاسترجاع من الإنترنت، من العناوين والاقتباسات (أي الأسطر القليلة الأولى في الموقع). والجمل غير المكتملة هي القاعدة لا الاستثناء في مثل هذه الاقتباسات.

وعادة ما تزود نظم استرجاع المعلومات المستفيد بوسائل مختلفة، للحصول على النتائج، وهي التحميل **download**، أو الطباعة، أو البريد الإلكتروني، أو العرض على الشاشة. وتدل كمية المخرجات على عدد النتائج التي يتم تقديمها في كل مرة يتقدم فيها المستفيد بطلب. وتكفل نظم الاسترجاع على الخط المباشر للمستفيد بوجه عام، تحديد أي عدد من النتائج التي تسترجع للمخرجات. وفي المقابل، تتطلب نظم الفهارس المتاحة على الخط المباشر، من المستفيدين، بيان خياراتهم، بتمييز أو اختيار التسجيلات كل على حدة أولا. ومن المعروف تماما أن نظم استرجاع الإنترنت، بإمكانها إنتاج الآلاف، إن لم يكن الملايين من نتائج عملية البحث، بسهولة. إلا أن المستفيد لا يمكنه عادة الحصول إلا على عشر نتائج أو عشرين، أو ثلاثين، في كل مرة يتقدم فيها بطلب للحصول على مخرجات. وليست هناك طريقة تكفل للمستفيد الحصول على جميع المخرجات على نحو فوري، إلا أن القصور الذي يكتنف

الحصول على نتائج البحث في الإنترنت، لا يشكل هما يذكر بالنسبة للمستفيد، نظرا لأن مخرجات نظم استرجاع الإنترنت، عادة ما ترتب طبقيا، وفقا لنظام تنازلي تبعا للصلاحية أو الارتباط بالموضوع، وقليل من المستفيدين من يرغب فعلا في الاطلاع على كل نتائج عملية البحث.

وخيارات المخرجات ليست بالعامل الحاسم في تفاعل المستفيد مع النظام، إلا أنه يمكن للمرونة في تقديم نتائج الاسترجاع، أن تدعم التعاطف مع المستفيد أو مراعاة ظروفه. وهناك تزايد في كفالة تصوير نتائج البحث بصرياً، من جانب بعض نظم استرجاع المعلومات؛ لمساعدة المستفيد في الحصول على نظرة تصويرية لما تم استرجاعه. وتبدو الفهارس المتاحة على الخط المباشر (مثل: **aqua.queenslibrary.org**) ونظم استرجاع الإنترنت (مثل: **www.quintura.com**) بوجه عام، سائرة بخطى أسرع من نظم الخط المباشر، في عرض نتائج البحث بصرياً. كذلك يوفر عدد من نظم استرجاع المعلومات صورا مصغرة لأغلفة الكتب (مثل: **books.google.com**) أو صفحات العنكبوتية (مثل: **www.ask.com**)، إلى جانب نتائج البحث النصية، حتى يكون لدى المستفيد المزيد من المعلومات للحكم على صلاحية النتائج المسترجعة.

مقومات المساعدة:

تؤثر مقومات المساعدة (التوجيهية، أو التعليمية، أو البطاقات الإرشادية، على سبيل المثال) إلى حد بعيد في مدى تعاطف نظم استرجاع المعلومات مع المستفيد، أو مراعاة ظروفه. وينبغي لمقومات المساعدة أن تساعد المستفيدين فعلا، في العثور على إجابات لما يمكن أن يكون لديهم من أسئلة، أثناء التفاعل مع نظم استرجاع المعلومات. وقد حدد ترنر **Trenner (1989)** معالم ستة من مثل هذه الأسئلة، وذلك

في دراسة تقارن مقومات المساعدة على الخط المباشر، في ستة عشر نظاما لاسترجاع المعلومات:

١. فيم أخطأت أنا، أو لما تلقيت رسالة الخطأ هذه؟

٢. هل أفعل أنا الشيء الصحيح؟

٣. أين أنا وماذا يمكن أن أفعل بعد ذلك؟

٤. كيف أفعل شيئا ما؟

٥. كيف أحصل على بعض المعلومات.

٦. لماذا فشل النظام؟

ويتعلق السؤال الأول بالموقف الذي يرتكبُ فيه المستفيد خطأ ما، أو يتلقى رسالة خطأ. ومن المفترض مثاليا، ألا تشير مقومات المساعدة إلى الخطأ الذي حدث فحسب، وإنما تحدد أيضا ما يمكن عمله لتصحيح الخطأ. إلا أن معظم مقومات المساعدة بنظم استرجاع المعلومات، قادرة على تنفيذ الأولى، ولكنها تعجز عن تنفيذ الأخيرة. ونتيجة لذلك تترك المستفيد يتخبط بلا هدي في حيرة مرتبكا. وفيما يتعلق ببقية الأسئلة التي سجلها ترنر **Trenner (1989)**، فإنه يمكن لمعظم مقومات المساعدة الإجابة عن السؤالين الرابع والخامس على نحو مناسب بشكل معقول، إلا أنها تبدو ضعيفة في التعامل مع الأسئلة الثاني، والثالث والسادس. وتبدو مقومات المساعدة، في الوقت الراهن، قادرة بوجه عام على تسجيل المعلومات الإجرائية الخاصة بأنشطة الاسترجاع. وعندما يتعلق الأمر بالأسئلة التي تتطلب توضيحا، فإن معظم مقومات المساعدة، في نظم استرجاع المعلومات تبدي عجزها.

وقد اقترح الباحثون (راجع على سبيل المثال **Shneiderman & Plaisant, 2005; Trenner, 1989**) إرشادات لتصميم مقومات المساعدة في النظم الإلكترونية، بما في ذلك نظم استرجاع المعلومات. ومن وجهة نظر المستفيد، فإن

مقومات المساعدة المناسبة ينبغي أن تكون قادرة على توفير الإجابات لأسئلة المستفيدين، وليس مجرد تقديم بعض البيانات التي يتم تزويد النظام بها سلفا. وينبغي أن تكون الإجابات أيضا خالية من المصطلحات المتخصصة، واللغة التي تنم عن التحدي أو التهديد (مثل "لقد تم تنفيذ عملية غير قانونية") وأن تكون هذه المقومات متاحة طوال الوقت، وللمستفيدين على اختلاف مستوياتهم؛ فالمستفيدون يلتمسون المساعدة من نظم استرجاع المعلومات عندما يحتاجون إليها، وإذا عجزت مقومات المساعدة عن تلبية احتياجات المستفيدين، فإن وظيفة نظم استرجاع المعلومات ذاتها قد لا تتحقق، وناهيك عن قضية التعاطف مع المستفيد أو مراعاة ظروفه.

تقييم تفاعل المستفيد مع النظام:

التعاطف مع المستفيد أو مراعاة ظروفه، معيار عريض لتقييم تفاعل المستفيد مع النظام. وقد تطورت مقاييس أكثر تحديدا (راجع على سبيل المثال **Shneidermann & Plaisant, 2005**) للهدف نفسه.

الوقت اللازم لتعلم المستفيد مهام استرجاع محددة:

يحدث التعلم حينما يتفاعل المستفيد مع نظام ما لاسترجاع المعلومات؛ أي أن المستفيدين بحاجة لأن يتعلموا كيف يتعاملون مع نظام ما لاسترجاع المعلومات، قبل إجراء عمليات البحث فعلا. وتنعكس سهولة التعلم على الوقت اللازم لاكتساب المستفيد لمهارة استرجاع عملية (كالبحث البوليني على سبيل المثال). وكلما كان وقت التعلم قصيراً، كان نظام استرجاع المعلومات أفضل، فيما يتعلق بتعاطفه مع المستفيد. ولا يوفر النظام غير الودي بيئة مريحة يمكن فيها للمستفيد أن يتعلم.

سرعة التفاعل:

كم يستغرق تلقي المستفيد للتلقيم المرتد من نظام استرجاع المعلومات، أثناء التفاعل؟ وكلما طال ما تستغرقه العملية من وقت، يصبح النظام أقل تعاطفا مع

المستفيد. وكما ذهبت شنايدرمان وبليزانت **Shneiderman and Plaisant (2005)**، فإن زمن الاستجابة بالنسبة للمهام المألوفة، ينبغي أن يكون ما بين ثانيتين وأربع ثوانٍ، ويمكن لوقت انتظار أطول من خمس عشرة ثانية أن يكون غير إيجابي بالمرة. وعلى الرغم من أن مؤشرات أزمنة الاستجابة هذه، تتعلق بالتفاعل بوجه عام، فإنها يمكن أن تنطبق ولاشك على بيئة استرجاع المعلومات. ويتقرر هذا المعيار، في الأساس، تبعاً للقدرة التجهيزية لنظام استرجاع المعلومات. وعادة ما تكفل وحدات التجهيز القوية سرعة تفاعل عالية.

معدل أخطاء المستفيد:

إذا كان نظام استرجاع المعلومات حدسيا، ويفسر نفسه ذاتيا، فإن المستفيدين يرتكبون قليلاً من الأخطاء في التفاعل معه. ومالم يكن الأمر كذلك فإن الأخطاء لا مفر منها (كاللجوء إلى خيار خاطئ على سبيل المثال). وبذلك يصبح عدد الأخطاء التي يقع فيها المستفيد، أثناء عملية الاسترجاع، مقياسا آخر للحكم على مدى تعاطف نظام استرجاع المعلومات مع المستفيد. ويمكن لمقومات المساعدة التي سبق أن ناقشناها، أن تنهض بدور مهم، في الحد من معدل الأخطاء. ومن الطبيعي أن يقع المستفيدون في أخطاء أقل، إذا ما توافرت المساعدة لهم، عندما تدعو الحاجة إليها، إذا ما كانت عملية التفاعل شفافة بالنسبة لهم.

القدرة على التذكر بمرور الوقت:

تستلزم إجراءات التفاعل، وخصوصا في أسلوب لغة إصدار الأوامر، أن يتم تعلمها والتمرس بها، لكي يتحقق التمكن منها. ويعد الاحتفاظ بهذه المهارات لفترة زمنية، تقييماً لمدى تعاطف نظم استرجاع المعلومات مع المستفيدين، من زاوية أخرى مختلفة؛ فمن الممكن على سبيل المثال، للغة إصدار الأوامر أن تكون من السهل

تذكرها، إذا كانت الأوامر قابلة للتذكر وتتسم بالاطراد، كأن يستعمل أمر الطبع كلمة **print** بدلا من **prt**، أو **type**، أو أي شيء آخر. ومن شأن واجهة التعامل المتعاطفة، نظرا للتعزيز الإيجابي، أن تمكن المستفيدين من الاحتفاظ بما يعرفونه حول النظام، لوقت أطول مما تكفله الواجهة غير المتعاطفة.

رضاء المستفيد:

وهذا المقياس لمدى تعاطف النظام ذاتي أو غير موضوعي؛ نظرا لأنه من الصعب تقييم الرضاء بطريقة موضوعية، حتى وإن كان من الممكن التعبير عن الرضاء كميا ببعض الطرق (مثل عدد النتائج المسترجعة)، إلا أن هذا المعيار ينبغي وضعه في الحسبان في تقييم الجانب النوعي في تفاعل المستفيد مع النظام؛ فكيف يمكن على سبيل المثال، أن يشعر المستفيد تجاه التفاعل؟ ويصنف بعض الباحثين (راجع **Hildereth, 2001** على سبيل المثال) رضاء المستفيد، إلى الرضاء المادي (أي النتائج التي يتم الحصول عليها من التفاعل مع نظام استرجاع المعلومات)، أو الرضاء العاطفي (أي الإحساس بالإنجاز الناتج عن التفاعل). وكلما كان التفاعل بوجه عام، أكثر تعاطفا، أصبح المستفيد أكثر رضاءً.

إلا أن هناك مواءمات لا مفر منها، بين المعايير الخمسة التي تناولناها توا (**Shneiderman &Plaisant, 2005**). ونادرا ما يبلغ أي تفاعل بين المستفيد والنظام قمة الأداء، في كل الجوانب. وينبغي دائما أن يوضع المستفيد في الحسبان؛ نظرا لأن جميع أنشطة استرجاع المعلومات تدور في فلك المستفيد، لا في فلك النظام.

وعلى الرغم من أنه لا يزال هناك الكثير مما ينبغي توافره في تفاعل المستفيد مع النظام، فإن الموقف قد تحسن ولاشك، إذ تم اكتساب المزيد من الخبرات في تلبية متطلبات المستفيد (**Large et al., 1999**)؛ فمن أسلوب لغة إصدار الأوامر، إلى

الأسلوب الهجين للتفاعل، ومن العرض أحادي اللون إلى التطبيق الفني لمختلف الخصائص، ومن رسائل المساعدة المحملة بالمصطلحات التقنية، إلى البطاقات الإرشادية الخالية من المصطلحات التخصصية، أصبح تفاعل المستفيد مع النظام أكثر تعاطفا مع المستفيد اليوم، عما كان عليه في الماضي.

المستفيد واسترجاع المعلومات في العصر الرقمي:

تؤدي التقنيات الرقمية ولاشك، إلى الارتفاع بمستوى كفاءة أنشطة استرجاع المعلومات. كذلك تثير هذه التقنيات بعض المخاوف والقضايا للمستفيد، وخصوصا المبتدئ، فاسترجاع المعلومات يعتمد على سبيل المثال، في العصر الرقمي، على التقنيات الرقمية، التي يمكن أن تسبب القلق والاغتراب للمستفيد، فقد لا يشعر كل مستفيد بالراحة عند إجراء عمليات البحث في نظام إلكتروني. ونظرا لما يكتنف النظم الراهنة لاسترجاع المعلومات من قصور، فإن المستفيد قد يشعر بالقلق أو عدم الاطمئنان عندما يحاول استرجاع المعلومات في البيئة الرقمية. يضاف إلى ذلك، الحاجة إلى تجهيزات متخصصة، لاسترجاع المعلومات رقميا. ولم يعد نهج القلم الرصاص والورق، في البحث عن المعلومات يدويا، في البيئة الورقية، يعمل بكفاءة في العصر الرقمي.

وفضلا عن ذلك، فإن نظم استرجاع المعلومات، في العصر الرقمي، يتم تطويرها باستمرار، أو يتم إدخال نظم جديدة. وقد يكون على المستفيد أن يتعلم أو يُدرَّب على أساس منتظم، لمتابعة التغيرات، التي تضيف أعباء جديدة لخطط عمل المستفيد المثقلة فعلا. وبعبارة أخرى، فإن استرجاع المعلومات في العصر الرقمي يمكن أن تكون له آثار جانبية بالنسبة للمستفيد؛ نظرا لعامل التقنيات.

وبالإضافة إلى عامل التقنيات، يتعين معاملة المستفيدين من نظم المعلومات المختلفة، على نحو مختلف؛ نظرا لأنهم عادة ما تكون لديهم خبرات استرجاع معلومات مختلفة؛ فالمستفيدون من نظم الخط المباشر، على سبيل المثال، إما أن يكونوا اختصاصيو معلومات، يقومون بدور الوسيط، أو مستفيدون نهائيون تلقوا تدريبا مناسبا. وفي مقابل ذلك، فإن المستفيدين من نظم استرجاع الإنترنت، قلما يكون قد توافر لهم التدريب في استرجاع المعلومات. ومن ثم، فإن بعد المستفيد أو ما يتعلق بالمستفيد يصبح متزايد التنوع بالنسبة لاسترجاع المعلومات في العصر الرقمي.

وعلى الرغم من أن تنظيم المعلومات في العصر الرقمي، قد لا يكون له ما لاسترجاع المعلومات من تأثير عن المستفيد، فإن تأثير التقنيات الرقمية في هذا المجال، لا يمكن تجاهله بحال. وما يجعل التأثير أقل وضوحا، هو أن من يقومون بتنظيم المعلومات، عادة ما يكونون اختصاصيي معلومات، تكيفوا مع إيقاع العصر الرقمي، وسرعته عن طريق تدريبهم وتمرسهم.

المراجع

Allen, Thomas. (1970). Communication networks in R & D laboratories. *R & D Management*, 1. Reprinted in Belver C. Griffith (Ed.) Key papers in information science (pp. 66-73). White plains, NY: Knowledge Industry Publications.

Bates, Marcia J. (1989). The design of browsing and berrypicking techniques for the online searching interface. *Online Review*, 13 (5), 407-424.

Belkin, Nicholas J., et al. (1995). Cases, scripts and information seeking strategies: On the design of interactive information retrieval systems. *Expert Systems With Applications*, 9(3), 379-395.

Belkin, Nicholas J., Oddy, Robert N., and Brooks, Helen M. (1982). ASK for information retrieval. Part I. Background and theory. *Journal of Documentation*, 38(2), 61-71.

Borgman, Christine L. (1989). All users of information retrieval systems are not created equal: An exploration into individual differences. *Information Processing & Management*, 25(3), 237-251.

Buente, Wayne, and Robbin, Alice. (2008). Trends in internet information behavior 2000-2004. *Journal of the American Society for Information Science and Technology*, 59(11), 1743-1760.

Byström, Katriina, and Järvelin, Kalervo. (1995). Task complexity affects information seeking and use. *Information Processing & Management*, 31(2), 191-213.

Case, Donald O. (2006). Information behavior. *Annual Review of Information Science and Technology*, 40, 293-327.

Case, Donald O. (2007). *Looking for information: A survey of research on information seeking, needs, and behavior.* 2nd ed. Boston: Academic Press.

Chowdhury, Gobinda G. (1999). *Introduction to modern information retrieval.* London: Library Association Publishing.

Co, Elise, and Pashenkov, Nikita. (2008). Emerging display technologies for organic user interfaces. *Communications of the ACM,* 51(6), 45-47.

Crane, Diane. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities.* Chicago: University of Chicago Press.

Dervin, Brenda (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology. In Jack D. Glazier and Ronald R. Powell (Eds.), *Qualitative Research in Information Management* (pp. 61-84). Englewood, CO: Libraries Unlimited.

Ellis, David. (1989). A behavioural approach to information retrieval design. *Journal of Documentation,* 45(3), 171-212.

Farooq, Umer, et al. (2007). Evaluating tagging behavior in social bookmarking systems: Metrics and design heuristics. *Proceedings of the ACM Conference on Supporting Group Work* (pp. 351-360). New York: Association of Computing Machinery.

Fenichel, Carol H. (1981). Online searching: Measures that discriminate among users with different types of experiences. *Journal of the American Society for Information Science,* 32, 23-32.

Frants, Valery I., and Brush, Craig B. (1988). The need for information and some aspects of information retrieval systems construction. *Journal of the American Society for Information Science,* 39(2), 86-91.

Galitz, Wilbert O. (1997). *Essential guide to user interface design: An introduction to GUI design: Principles and techniques.* New York: John Wiley & Sons.

Gorman, Paul, et al. (2002). Following experts at work in their own information spaces: Using observational methods to develop tools for the digital library. *Journal of the American Society for Information Science and Technology,* 53(14), 1245-1250.

Hildreth, Charles. (2001). Accounting for users' inflated assessments of online catalog search performance and usefulness: An experimental study. *Information Research,* 6(2). Retrieved January 21, 2009, from informationr.net/ir/6-2/paperl01.html

Huotari, Maija-Leena, and Chatman, Elfreda. (2001). Using everyday life information seeking to explain organizational behavior. *Library & Information Science Research,* 23(4), 351-366.

Ingwersen, Peter. (1982). Search procedures in the library analyzed from the cognitive point *of view. Journal of Documentation,* 38,165-191.

Ingwersen, Peter. (1992). *Information retrieval interaction.* London: Taylor Graham.

Ingwersen, Peter. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation,* 52(1), 3-50.

Ingwersen, Peter, and Järvelin, Kalervo. (2005). *The turn: Integrating of information seeking and retrieval in context.* Dordrecht, The Netherlands: Springer.

Järvelin, Kalervo, and Wilson, Tom D. (2003). On conceptual models for information seeking and retrieval research. *Information Research,* 9(1), paper 163. Retrieved December 20, 2008, from informationr.net/ir/9-l/ paperl63.html

Julien, Heidi, and Duggan, Lawrence J. (2000). A longitudinal analysis of the information needs and uses literature. *Library & Information Science Research,* 22 (3), 291-309.

Kipp, Margaret E. I., and Campbell, D. Grant. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. *Proceedings of the Annual Meeting of the American Society for Information Science and Technology,* 43. [CD-ROM]. Retrieved January 5, 2009, from dlist.sir.arizona.edu/1704/ 01/KippCampbellASIST.pdf.

Komlodi, Anita, Soergel, Dagobert, and Marchionini, Gary. (2006). Search histories for user support in user interfaces. *Journal of the American Society for Information Science and Technology,* 57(6), 803-807.

Kuhlthau, Carol C. (1993). *Seeking information: A process approach to library and information services.* Norwood, NJ: Alblex.

Lancaster, F. W., and Warner, Amy J. (1993). *Information retrieval today.* Arlington, VA: Information Resources Press.

Large, Andrew, Tedd, Lucy A., and Hartley, R. J. (1999). *Information seeking in the online age: Principles and practice.* London: Bowker-Saur.

Marchionini, Gary, et al. (1993). Information seeking in full-text end-user-oriented search systems: The roles of domain and search expertise. *Library & Information Science Research,* 15,35-69.

Miller, George A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 63, 81-97.

Paisley, William J. (1968). Information needs and uses. *Annual Review of Information Science and Technology,* 3,1-30.

Parkes, Amanda, Poupyrev, Ivan, and Ishii, Hiroshi. (2008). Designing kinetic interactions for organic user interfaces. *Communications of the ACM,* 51(6), 58-65.

Pettigrew, Karen E., Fidel, Raya, and Bruce, Harry. (2001). Conceptual frameworks in information behavior. *Annual Review of Information Science and Technology,* 35, 43-78.

Pirolli, Peter, and Card, Stuart K. (1999). Information foraging. *Psychological Review,* 106(4), 643-675.

Price, Derek J. de Solla. (1963). *Little science, big science.* New York: Columbia University Press.

Rekimoto, Jun. (2008). Organic interaction technologies: From stone to skin. *Communications of the ACM,* 51(6), 38-44.

Resnick, Marc L., and Vaughan, Misha W. (2006). Best interface and future visions for search user interfaces. *Journal of the American Society for Information Science and Technology,* 57(6), 781-787.

Rogers, Yvonne. (2004). New theoretical approaches for human-computer interaction. *Annual Review of Information Science and Technology,* 38, 87-143.

Rose, Daniel E. (2006). Reconciling information-seeking behavior with search user interfaces for the web. *Journal of the American Society for Information Science and Technology,* 57(6), 797-799.

Rowlands, lan, et al. (2008). The Google generation: The information behaviour of the researcher of the future. *ASLIB Proceedings,* 60(4), 290-310.

Shaw, Debora. (1991). The human-computer interface for information retrieval. *Annual Review of Information Science and Technology,* 26, 155-195.

Shneiderman, Ben, and Plaisant, Catherine. (2005). *Designing the user interface: Strategies for effective human-computer interaction.* 4th ed. Boston: Addison-Wesley.

Soukhanov, Anne H., et al. (Eds.). (1984). *Webster's II new Riverside university dictionary.* Boston: Riverside Publishing Co.

Spink, Amanda, and Cole, Charles. (2006). Human information behavior: Integrating diverse approaches and information use. *Journal of the American Society for Information Science and Technology,* 57(1), 25-35.

Trenner, L. (1989). A comparative survey of the friendliness of online "help" in interactive information retrieval systems. *Information Processing & Management,* 25(2), 119-136.

Vertegaal, Roel, and Poupyrev, Ivan. (2008). Organic user interfaces. *Communications of the ACM,* 51(6), 26-30.

Wilson, Tom D. (1999). Models of information behaviour research. *Journal of Documentation,* 55(3), 249-270.

Wilson, Tom D. (2008). The information user: Past, present and future. *Journal of Information Science,* 34(4), 457-464.

Zhang, Xiangmin, and Chignell, Mark. (2001). Assessment of the effects of user characteristics on mental models of information retrieval systems. *Journal of the American Society for Information Science and Technology,* 52(6), 445-459.

الفصل الحادي عشر

تقييم تنظيم المعلومات واسترجاعها

الآن، وقد ناقشنا مختلف جوانب ومكونات تنظيم المعلومات واسترجاعها، حان الوقت لتقييم هذه الأنشطة المترابطة. وقد فصَّل كثير من الباحثين (راجع، **Keen, 1971; Large, Tedd & Hartley, 1999; Swanson R.W., 1978** على سبيل المثال) القول في الحاجة إلى تقييم تنظيم المعلومات واسترجاعها. ونأمل في الارتقاء بمستوى الأداء في هذه الأنشطة، عندما يتم التحقق من مظاهر ضعف نظم استرجاع المعلومات، ثم التخلص من هذه المظاهر في النهاية، نتيجة لعملية التقييم.

وتقييم تنظيم المعلومات واسترجاعها، موضوع يحظى باهتمام كثير جدا من الباحثين في المجال، وذلك على مدى أكثر من خمسين عاما، إلا أن هناك كثيرا من الجدل حول مقاييس التقييم وطرقه. وسوف نناقش في هذا الفصل ثلاثة موضوعات رئيسة: (١) مقاييس التقييم (كالتحقيق **precision** والاستدعاء **recall** على سبيل المثال) لأداء تنظيم المعلومات واسترجاعها، (٢) معايير التقييم بالنسبة لمختلف أنواع نظم استرجاع المعلومات، بما في ذلك نبذة عن القابلية للإفادة، (٣) مشروعان كبيران لتقييم تنظيم المعلومات واسترجاعها، هما اختبارات أو تجارب كرانفيلد **Cranfield**، وسلسلة مؤتمرات استرجاع النصوص (ترك **Trec**) **Text REtrieval Conference**. ونتناول قضية طرق التقييم، بالإضافة إلى التجارب الكبرى في المجال.

مقاييس تقييم تنظيم المعلومات واسترجاعها:

تطورت مقاييس كثيرة لتقييم تنظيم المعلومات واسترجاعها. ومن أجل الوضوح نناقش مقاييس تقييم تنظيم المعلومات، بمعزل عن مقاييس تقييم استرجاع المعلومات.

مقاييس تقييم تنظيم المعلومات:

كما بينا في الفصلين الثاني والثالث، يمكن للمعلومات أن تنظم بطرق متنوعة. والتكشيف، والتقسيم إلى فئات، أي التصنيف، والتلخيص، بعض الأشكال الرئيسة لتنظيم المعلومات. وعلى الرغم من أن لكل شكل من أشكال تنظيم المعلومات خصائصه التي تميزه، فإنه قد قصد بكل هذه الأشكال مجتمعة، إبراز جوهر المعلومات، وجعل المعلومات قابلة للاسترجاع فيما بعد، ومن ثم فإن مجموعة مشتركة من معايير التقييم (كالدقة، والإيجاز، والاطراد، والموضوعية)، يمكن أن تطبق على كل أنواع طرق تنظيم المعلومات.

الدقة:

الدقة **accuracy** هي أهم مقياس في تقييم جودة تنظيم المعلومات؛ نظرا لأن التنظيم المشوه للوثائق الأصلية، يمكن أن يضيع المغزى الكامل لتنظيمها. وتعني الدقة أن يمثل التنظيم محتوى الوثائق الأصلية بدقة. ويمكن للاختزال في التنظيم أن يحدث فيما يتعلق بالحجم، وليس فيما يتعلق بالمحتوى أو المضمون، وينبغي وضع الدقة في الحسبان أيضا، في الاستشهاد المرجعي **citation** والهجاء على سبيل المثال.

والتحديد أو التخصيص **specificity** مفهوم وثيق الصلة بالدقة، وعادة ما يستخدم لقياس جودة التكشيف. ويعني التحديد، بحكم التعريف، المدى الذي يتم به التحقق من مفهوم ما أو موضوع ما في وثيقة ما، بمصطلح دقيق في سلم أو هرم

علاقات الجنس والنوع الخاصة به (Wellisch, 1995). ففي هرم "استرجاع المعلومات" - "استرجاع الوسائط المتعددة" - "استرجاع الفيديو" information retrieval-multimedia retrieval-video retrieval، يتم تعيين المصطلح "استرجاع الفيديو" للوثيقة التي تتناول الموضوع، لضمان أعلى درجات التخصيص في التكشيف. أما المصطلحان الآخران في الهرم، فأدنى درجة في التخصيص. ويتحدد التخصيص بلغة التكشيف (أي اللغة المقيدة أو اللغة الطبيعية) المستخدمة. وتكفل اللغة الطبيعية للمرء بوجه عام، القدرة على تحقيق درجة عالية من التخصيص في التكشيف.

الإيجاز:

من بين أهداف تنظيم المعلومات، مساعدة المستفيدين على الاقتصاد في الوقت والجهد، في الحصول على ما يحتاجون إليه من معلومات. وتفجر المعلومات إحدى العبارات التي تم سكها لوصف الكميات الهائلة من المعلومات، التي يتعين على المستفيدين التعامل معها. ومن ناحية أخرى، فإنه من الحقائق التي لا يمكن إنكارها، أن الأمر يتطلب من المستفيد، لمراجعة ناتج تنظيم المعلومات (كالمستخلصات، والملخصات، ومصطلحات التكشيف على سبيل المثال) وقتا أقصر بكثير مما يتطلبه الاطلاع على وثيقة أصلية؛ ومن ثم فإن الإيجاز يصبح مقياسا رئيسا، للحكم على ما إذا كان هدف الاقتصاد في الوقت والجهد قد تحقق فعلا أم لا.

وقد تم إقرار معايير معينة لمختلف أشكال تنظيم المعلومات (راجع، Lancaster, 1998 على سبيل المثال) فيما يتعلق بالإيجاز؛ فبالنسبة لمقالات الدوريات، على سبيل المثال، عادة ما يتم تعيين ما بين خمسة مصطلحات تكشيف وعشرة، وإعداد مستخلص من ٢٥٠ كلمة أو أقل. وإذا ما كان ناتج التنظيم يتخذ شكلاً مطولاً، فإنه قلما يكفل مساعدة المستفيدين على التذكر، من أجل الحد من وقتهم

المستنفد في استرجاع المعلومات. ومن الممكن معاملة الشمول **exhaustivity** في التكشيف بوصفه مصطلحا يتصل بالإيجاز. وقد عرفت ولّش **Wellisch (1995)** الشمول بوصفه المدى الذي يتم به جعل المفاهيم والموضوعات قابلة للاسترجاع، بواسطة مصطلحات التكشيف. ويتحدد الشمول بناء على سياسة التكشيف المتبعة في النظام.

الاطـراد:

من الممكن معالجة الوثيقة الواحدة بعدة أشكال (كالتكشيف والاستخلاص على سبيل المثال) ومن جانب عدة أشخاص. كما أنه من المألوف أيضا أن نرى مطبوعا ما (دورية على سبيل المثال) مكشفا من جانب خدمات تكشيف مختلفة؛ فمجلة **Journal of the American Society for Information Science and Technology**، على سبيل المثال، تكشّف في **Chemical Abstracts**، و **Computer Literature Index**، وعشرة مراصد أخرى لخدمات الاستخلاص والتكشيف. وعندما يطبق هذا الأسلوب للمعالجة المتعددة للمعلومات، الذي اقترحه إنجورسن **Ingwersen (1996)**، فإن الناتج النهائي يمكن أن يكون غير مطرد في التغطية والأسلوب. وقد تبين من دراسات التكشيف، أن المستوى المرتفع للاطراد **consistency** من الصعب بلوغه، وأن اطراد التكشيف يتراوح بين ١٠ بالمئة و ٨٠ بالمئة (راجع على سبيل المثال **Hooper, 1965**). وقد سجل كلفردون **Cleverdon (1984)** أنه عندما كشف مكشفان متمرسان الوثيقة نفسها، باستخدام اللغة المقيدة نفسها، كانت نسبة المصطلحات المشتركة فيما بينهما ٣٠ بالمئة فقط. ومن الممكن لتنظيم المعلومات، في العصر الرقمي، بفضل الأساليب الآلية، أن يحقق نظريا اطرادا أفضل، إلا أن البحوث التجريبية في هذا المجال، لم يتم إجراؤها بعد.

ولا يعني الاطراد ضمنا مجرد مضاهاة كلمة بكلمة، وإنما يدل على المعالجة الموحدة (كاللغة وعمق التغطية على سبيل المثال) لجوهر الوثائق الأصلية، بصرف النظر عن شكل التنظيم، والشخص الذي يقوم بالتنظيم. ويمكن للاطراد أن يؤثر في جودة التنظيم، وأداء الاسترجاع في النهاية أيضا.

الموضوعية:

تشكل الموضوعية **objectivity** معيارا آخر للتقييم بالنسبة لتنظيم المعلومات؛ نظرا لأن الذاتية **subjectivity**، والتفسيرات الشخصية، يمكن أن تظهر في أثناء عملية التنظيم. وينبغي لتنظيم المعلومات أن يكون توصيفا موثوقا به للوثيقة الأصلية، بدون أي تغيير في محتواها. ومن المهم أيضا ألا يتم إسباغ أو إضفاء الأفضليات الشخصية للمرء على التنظيم. وبمزيد من التحديد، فإن أي تعزيز لأفكار المؤلف، أو الإقحام المتعمد لاتجاهات المكشف، أو اختيار مصطلحات تعبر ضمنا عن الاتجاهات، أو تجاهل أو إهمال أفكار وردت بالوثيقة، في الكشاف، ينال فعلاً من موضوعية تنظيم المعلومات (**Bell, 1991**). فالتنظيم المتسم بالذاتية لا يعبر عن الصورة الحقيقية للوثائق الأصلية.

الوضوح، والقابلية للقراءة، والقابلية للإفادة:

بالإضافة إلى المقاييس التي تناولناها توا، طرحت معايير أخرى لتقييم تنظيم المعلومات. والوضوح **clarity**، والقابلية للقراءة أو الانقرائية **readability**، والقابلية للإفادة أو الاستخدام **usability**، من بين هذه المعايير. وتحاول كل هذه المقاييس إبداء التعاطف مع المستفيد أو تيسير مهمته **user friendliness**، وجعل المستفيدين يشعرون بالراحة أكثر، عندما يتعاملون مع ناتج تنظيم المعلومات.

ومقاييس التقييم بالنسبة لتنظيم المعلومات، أقل من تلك الخاصة باسترجاع المعلومات، كما أن هذا الضرب من التقييم لا يميز أيضا عملية التنظيم عن أداء التنظيم. وإذا أردنا الدقة، فإن جودة التنظيم لا يمكن أن تقيم فعلا، ما لم يستعمل ناتج التنظيم في استرجاع المعلومات. وفي الوقت نفسه، تتوقف جودة استرجاع المعلومات على جودة تنظيم المعلومات. فالتنظيم الجيد للمعلومات، ينهض بدور لا غنى عنه في ضمان أداء الاسترجاع الجيد، ولا يمكن إلا من خلال استرجاع المعلومات الحكم في النهاية على جودة تنظيم المعلومات.

مقاييس تقييم استرجاع المعلومات:

بالمقارنة بتنظيم المعلومات، حظى استرجاع المعلومات بقدر كبير جدا من الاهتمام، من جانب الباحثين، فيما يتعلق بمقاييس التقييم، ويرجع ذلك جزئيا إلى أن الأول مجرد شرط أساس للأخير. وكما سبق أن ذكرنا، فإن جودة التنظيم لا يمكن تقييمها فعلا إلا عندما يستعمل ناتج التنظيم في الاسترجاع. واسترجاع المعلومات نشاط يركز على الأداء، كما يركز على العملية في الوقت نفسه، على الرغم من أن البشر، وخصوصا المستفيدين النهائيين، عادة ما يبدون اهتماما بما يتم استرجاعه، أكثر مما يهتمون بما يتم فعلا أثناء عملية الاسترجاع. إلا أنه يمكن للفهم الأفضل لعملية الاسترجاع، أن يؤدي إلى الارتفاع بمستوى أداء الاسترجاع. ولهذا السبب، فإن مقاييس التقييم التي تطورت لاسترجاع المعلومات، يمكن تقسيمها تقريبا إلى نوعين؛ أولهما يركز على الأداء، ويركز الآخر على العملية. ونناقش النوع الأول من المقاييس في ١١/ ١/ ٢، بينما تحظى مقاييس تقييم عملية الاسترجاع، أو نظام استرجاع المعلومات برمته، بالاهتمام في ١١/ ٢. ومن بين مقاييس تقييم أداء الاسترجاع، يعد الاستدعاء والتحقيق أشهر معيارين، وهما الأوسع مدى في الاستخدام **(Salton, 1992)**.

الاستدعاء والتحقيق:

يقدم الجدول رقم ١١/ ١ جميع النتائج المحتملة بالنسبة لمهمة الاسترجاع، باستعمال الرموز والمصطلحات التي عادة ما نجدها في الوثائق التي تتناول استرجاع المعلومات. فالتسديدات "**hits**" على سبيل المثال، تعني الوثائق المسترجعة الصالحة أو المرتبطة بالموضوع **relevant**. وتعرف الاخفاقات "**misses**" بأنها الوثائق الصالحة التي لم تسترجع. وبناء على نتائج الاسترجاع المحتملة هذه، أمكن اشتقاق مقاييس تقييم متنوعة. والاستدعاء والتحقيق، اللذان اقترحهما في البداية كنت وزملاؤه **Kent and his colleagues (1955)**، كمعامل الصلاحية **pertinency factor**، ومعامل الاستدعاء **recall factor**، اثنان من مثل هذه المقاييس.

الجدول رقم ١١/ ١ نتائج الاسترجاع المحتملة

الحكم / النتيجة	متصل بالموضوع	غير متصل بالموضوع	المجموع
مسترجع	أ (التسديدات)	ب (الشوشرة)	أ + ب (كلها مسترجعة *)
غير مسترجع	ج (الإخفاقات)	د (المرفوضة)	ج + د (كلها غير مسترجعة)
المجموع	أ + جـ (كلها متصلة بالموضوع)	ب + د (كلها غير متصلة بالموضوع)	أ + ب + ج + د (كل ما في النظام)

وكما تبين لنا في ٥/ ٢/ ٢، فإن الاستدعاء (R في المعادلات التالية) يعرف بوصفه نسبة عدد الوثائق المتصلة بالموضوع المسترجعة، إلى إجمالى عدد الوثائق المتصلة بالموضوع في النظام. وباستعمال الرموز الواردة في الجدول رقم ١١/ ١، يمكن صياغة المعادلة التالية:

R = أ / (أ + جـ)

وبناء على ذلك فإن التحقيق (**P**) هو نسبة عدد الوثائق المتصلة بالموضوع المسترجعة، إلى إجمالى عدد الوثائق المسترجعة من النظام:

P = أ/ (أ + ب)

وتميل العلاقة بين الاستدعاء والتحقيق لأن تكون تناسبا عكسيا (راجع على سبيل المثال **Lancaster & Warner, 1993**)، وإن كان فاجمان **Fugmann (1993)** قد دحض تلك العبارة بعدة أمثلة؛ فقد تبين لفاجمان أن الزيادة في التحقيق ليس من الضروري أن تكون مصحوبة بانخفاض مقابل في الاستدعاء، وأن الزيادة في الاستدعاء لا يمكن بحال القول بأنها يترتب عليها دائما تناقص في التحقيق. والتفسير الجزئي للنتيجة التي انتهى إليها فاجمان، هو أن استرجاع الوثائق الصالحة يؤثر في كل من الاستدعاء والتحقيق، بينما يؤثر استرجاع الوثائق غير الصالحة في التحقيق فقط **(Korfhage, 1997)**.

ومن المفترض أن يقيس الاستدعاء القابلية للاسترجاع الخاصة بنظام استرجاع المعلومات، بينما ينبغي أن يقيم التحقيق قدرة نظام استرجاع المعلومات على فصل غير الصالح عن الصالح. إلا أن هناك عقبتين رئيستين تمثلان حجر عثرة في حساب هذين المقياسين؛ أولاهما، كيف يمكن تعريف الصلاحية وقياسها؟ والثانية كيف يمكن التعرف على إجمالى عدد الوثائق الصالحة في النظام؟

مفهوم الصلاحية:

لقد كان النقاش، وبمزيد من الدقة الجدل، حول معنى الصلاحية طويلا ومحتدما. وقد قدم ميزارو **Mizzaro (1997)** تاريخًا للصلاحية، بمراجعة حوالي ١٦٠ بحثا، نشرت عن الموضوع، منذ ما قبل خمسينيات القرن العشرين، حتى منتصف تسعينيات القرن نفسه. وعلى الرغم من ذلك حظي الأمر بقدر كبير من الاهتمام، ولم يتم الاتفاق حول سبل تعريف الصلاحية، وكيف يمكن قياسها ببساطة؛ نظرا لأن الصلاحية، في جوهرها حكم ذاتي، متأثر بكثير من العوامل المترابطة؛ فالصلاحية كالحب، إحساس في نظر من يرى. وهي مفهوم لا يمكن وصفه صراحة أو التعبير عنه كميا بسهولة.

وعلى الرغم من ذلك، تم اقتراح عدد ضخم من التعريفات للصلاحية. وقد تتبع ساراسفك **Saracevic (1975)**، وحلل ولخص كثيرا من التعريفات المتوافرة للصلاحية، وخلص إلى أن فهمنا للصلاحية قد أصبح أفضل بكثير، وأوضح وأعمق وأوسع مما كان عليه في أعقاب الحرب العالمية الثانية مباشرة. إلا أن الشوط لا يزال طويلا. وقد أعاد شامبر، وآيزنبرج، ونيلان **Schamber, Eisenberg and Nilan (1990)** النظر في قضية تعريف الصلاحية، وأشاروا إلى أنه لم يتم التوصل بعد إلى اتفاق فيما يتعلق بالتعريف. وكان في اعتقاد هؤلاء الباحثين، أن الصلاحية مفهوم متعدد الأبعاد، يتوقف على عوامل داخلية وخارجية. كما يستند هذا المفهوم أيضا إلى عملية إبداء رأي بشرية ديناميكية. ومن الممكن، على وجه التحديد، التحقق من ثمانين عاملا مؤثرا في أحكام الصلاحية، وتصنف هذه العوامل في ست فئات: من يصدرون الأحكام، والطلبات، والوثائق، ونظم استرجاع المعلومات، وظروف إصدار الأحكام، واختيار المقياس **(Schamber, 1994)**. وفضلا عن ذلك، يؤثر عامل

التعلم، وتسلسل عرض نتائج عملية البحث أيضا في أحكام الصلاحية (Froehlich, 1994). فبإمكان من يصدر الحكم أن يتعلم في أثناء عملية الحكم على الصلاحية. ومن المحتمل للوثيقة نفسها أن يصدر الحكم بأنها أكثر صلاحية، إذا ما قدمت في موقع متقدم لا في موقع متأخر في عملية التقييم. وغالبا ما يضيف إلى التعقد المتزايد فعلا، والصعوبة والإحباط في الحكم على الصلاحية، قيام شخص ما، خلاف المستفيد النهائي (كاختصاصي المعلومات) بالحكم على الصلاحية، وخصوصا في بيئات عمليات البحث التي تحظى بالمساعدة، كما في نظم الخط المباشر.

فالصلاحية مفهوم معقد فعلا، إلا أنها ظاهرة نسقية وقابلة للقياس (Schamber, Eisenberg & Nilan, 1990). ويمكن القول ببساطة أن الصلاحية خاصية تعبر عن العلاقة بين الوثيقة واستفسار المستفيد. وبينما يبدو من الصعب إصدار حكم دقيق وموضوعي على الصلاحية، فإنه من الممكن الحصول على تقدير لقيمتها. ويتعين على من يقومون بالتقييم، أن يضعوا في الوقت نفسه في الحسبان، القضايا التي طرحت ونوقشت في هذا القسم الفرعي، عندما يكونون بصدد الحكم على صلاحية نتائج الاسترجاع.

وبعد أكثر من ثلاثين عاما من نشر مراجعته العلمية الضافية حول الصلاحية (Saracevic, 1975)، أعد ساراسفك مراجعة علمية من جزأين بالعنوان نفسه، إلا أنه ميزهما بالجزء الثاني (Saracevic, 2007a) والجزء الثالث (Saracevic, 2007b)، ليبين أنهما معا، استطراد لعمله البذري الذي نشر عام ١٩٧٥، وإن لم يكن حينئذ قد عنون بالجزء الأول. وبعد المزيد من المراجعات العلمية لمفهوم الصلاحية، من حيث طبيعته، وتجلياته، ومضامينه، خلص ساراسفك إلى أن الصلاحية، بوصفها المفهوم الأساس في استرجاع المعلومات، جوهرية؛ ومن ثم فإن البحث والنقاش حول الصلاحية سيظل دائما في وقته وأوانه.

تحديد جميع الوثائق الصالحة في نظام ما:

طريقة تعريف الصلاحية وقياسها مجرد مشكلة واحدة فقط، في حساب الاستدعاء والتحقيق. وينطوي المأزق الآخر على تحديد كل الوثائق الصالحة لاستفسار ما في نظام ما لاسترجاع المعلومات. وما لم يكن النظام قد أنشئ لأغراض تجريبية، فإنه يبدو أن لا سبيل لمعرفة مجموع الوثائق الصالحة في النظام، لاستفسار معين على وجه التحديد. إلا أن ذلك العدد، المقام في المعادلة:

R = أ/ (أ+ جـ)، لا غنى عنه بالنسبة لحساب الاستدعاء.

وقد اقترحت عدة طرق، على مر السنين، للتغلب على هذه المشكلة **(Large, Tedd & Hartley, 1999)**. والتجمع **pooling** واحد من مثل هذه الطرق؛ ففي المقام الأول يقوم مستفيدون مختلفون بإجراء عملية البحث نفسها، على افتراض أنه إذا ما تمت عدة محاولات كافية لإجراء عملية البحث، فإن معظم، إن لم يكن كل الوثائق الصالحة يمكن العثور عليها في النهاية مرة واحدة على الأقل، ثم يتم بعد ذلك، تجميع النتائج التي أسفرت عنها جميع عمليات البحث معا، مع استبعاد المكررات. ولحساب الاستدعاء حينئذ، يمكن مقارنة نتائج أي من عمليات البحث، مقابل هذه الفئة المجمعة من الوثائق الصالحة في النظام. وقد تم تطبيق طريقة التجمع هذه في سلسلة مؤتمرات استرجاع النصوص **TREC**، لحساب الاستدعاء **(Harman, 1995)**، بالإضافة إلى مشروعات أخرى (راجع على سبيل المثال **Clarke & Willett, 1997**). ومن ثم، يثار السؤال حول كم عدد عمليات البحث التكرارية يتطلب الأمر إجراؤها، حتى يشعر أولئك الذين يستخدمون طريقة التجمع بالثقة بأنه قد تم تجميع ما يكفي من الوثائق الصالحة.

والنهج الآخر لتقدير إجمالي عدد الوثائق [الصالحة] في نظام ما، هو أن يطلب من خبير واحد أو أكثر، تجمعهم خبرة إجراء عمليات البحث والمعرفة الموضوعية،

إجراء عمليات البحث، مع الافتراض بأن هؤلاء الخبراء سوف يعثرون على جميع الوثائق الصالحة في النظام. ومن الممكن مراجعة النتائج التي أمكن الحصول عليها من جانب فئة أخرى ممن يجرون عمليات البحث، كفئة المبتدئين، مقابل عمليات بحث هؤلاء الخبراء؛ لقياس الاستدعاء. ويمكن تسمية هذا النهج بطريقة الخبراء.

وبالإضافة إلى طريقة التجمع، وطريقة الخبراء، تناول كل من لارج وتد وهارتلي **Large, Tedd & Hartley (1999)** طريقة الفئة الفرعية **subset**. وفي هذه الطريقة، يتم فحص الوثائق الواردة في فئة فرعية صغيرة من مرصد بيانات النظام، كلا على حدة، للعثور على الوثائق الصالحة. والافتراض الذي يُستند إليه هنا هو أن المجموعة الفرعية أو الفئة الفرعية **subset** تمثل النظام برمته، مما يوحي بضرورة تطبيق أسلوب العينة العشوائية، لاختيار الفئة أو المجموعة الفرعية. وحينئذ يصبح التقدير القائم على الفئة الفرعية هو المقام لحساب الاستدعاء.

وكما هو واضح، فإن الحصول على مقياس للاستدعاء لتقييم أداء الاسترجاع، أشبه بإحصاء عدد الأسماك في البحيرة. والمهمة أقرب للمستحيل؛ نظرا لأن مرصد بيانات النظام ديناميكي، دائم التغير. وجميع الطرق التي ناقشناها هنا، إنما هي في أحسن الأحوال تقديرات جيدة، وفي أسوأ الأحوال محصن تخمينات صريحة **(Large, Tedd & Hartley, 1999)**.

انتقادات الاستدعاء والتحقيق الأخرى:

لم يتم التوصل، كما يبدو، إلى إجابات مرضية كافية للسؤالين اللذين سبق طرحهما، بشأن الصلاحية، ومجموع عدد الوثائق الصالحه في نظام ما، إلا أن مزيدا من الانتقادات وجهت للاستدعاء والتحقيق كمقياسين للتقييم؛ فقد تمت الإشارة، على سبيل المثال، إلى أن الاستدعاء والتحقيق مقياسان غير كاملين للتقييم؛ نظرا لأن

عوامل كالتفاعلية، والتكلفة، وسرعة الاسترجاع، ينبغي أيضا أن توضع في الحسبان (راجع على سبيل المثال **Large, Tedd & Hartley, 1999**). كذلك ذهب ملر وتجلر **Miller and Tegler (1986)** إلى أن:

> الطرق التقليدية لتقييم عمليات البحث عن المعلومات - الاستدعاء والتحقيق - قد أغفلت تماما هذا الجانب التوليدي الإبداعي لعملية البحث. وبتقييم الناتج وليس العملية، فإن الاستدعاء والتحقيق يحدان من فهمنا لعمليات البحث عن المعلومات، ويعجزان عن قياس هذه العمليات على نحو فعال. **(p. 371)**

وبعبارة أخرى، فإن مقاييس أخرى خلاف الاستدعاء والتحقيق، ينبغي أن تطبق في تقييم استرجاع المعلومات. وكما أشار سالتون **Salton (1992)**، فإن الافتراض الرئيس وراء استخدام مقاييس كالاستدعاء والتحقيق، هو أن المستفيد العادي يهتم باسترجاع كميات كبيرة من المواد الصالحة (مما يسفر عن استدعاء مرتفع) بينما يرفض في الوقت نفسه نسبة كبيرة من المواد الهامشية (مما يسفر عن تحقيق مرتفع). وقد لا تتحقق هذه الافتراضات على الدوام؛ ففيما عدا حالات كاسترجاع براءات الاختراع، على سبيل المثال، فإنه قد لا يكون المستفيدون حريصين على عمليات بحث شاملة، الأمر الذي يشكك في صلاحية الاستدعاء كمقياس للتقييم.

يضاف إلى ذلك، أنه ليست هناك طريقة للتعبير عن الدرجات المختلفة للصلاحية عند حساب الاستدعاء؛ فجميع الوثائق الصالحة يفترض عادة أن تكون ذات قيمة متساوية، بصرف النظر عما إذا كانت صلاحيتها هامشية أو بؤرية **(Chowdhury, 1999)**. ولا يمكن تغيير هذا الافتراض، وخصوصا عندما يكون

العدد الإجمالي للوثائق الصالحة في نظام ما، بحاجة إلى تقدير بواسطة الطرق التي تناولناها في ١١/ ١/ ٢/ ١/ ٢.

ويبدو الاستدعاء والتحقيق كمقياسين للتقييم، أبعد ما يكونان عن المثالية، كما يبدو الاستدعاء أكثر المقياسين إثارة للجدل. إلا أنه على الرغم من كل الانتقادات والجدل، فإن مقياسي الاستدعاء والتحقيق يشكلان الأساس بالنسبة لعمليات التقييم المعروفة، لكل من نظم الاسترجاع العاملة والمختبرية أو التجريبية **(Salton, 1992)**. وقد تطورت في الوقت نفسه أشكال متعددة لمقياسي الاستدعاء والتحقيق، بهدف التخلص مما يكتنف معياري التقييم هذين من أوجه القصور.

الأشكال المختلفة لمقياسي الاستدعاء والتحقيق:

لقد تم اقتراح متوسط الاستدعاء أو التحقيق، لدحض افتراض القيمة المتساوية لصلاحية الوثائق المسترجعة. ولم تعد الصلاحية في هذه الطريقة خاصية ثنائية (أي صالح أو غير صالح) في حساب متوسط الاستدعاء أو التحقيق. وعوضا عن ذلك فإن الصلاحية تقاس بمقياس من س من النقاط (س= ٣، ٤، ...١١) لبيان مختلف درجات الصلاحية. وإذا ما استعمل مقياس من ثلاث نقاط على سبيل المثال، فإنه يمكن إعطاء قيمة من صفر أو ٠.٥ أو ١.٠ لكل وثيقة مسترجعة. ومن الممكن بعد ذلك حساب متوسط القيم المعطاة لجميع الوثائق المسترجعة حول موضوع ما، بحيث يمكن وضع تفاوت الصلاحية في الوثائق المسترجعة، في الحسبان. وليس من الضروري أن يقتصر عدد النقاط في المقياس، على ١١ أو أقل، إلا أنه يمكن للتقسيم الأكثر دقة أن يعقّد عملية الحساب. وكما سبق أن بينا، فإن حساب متوسط الاستدعاء، أصعب من قياس متوسط التحقيق؛ نظرا لصعوبة تحديد العدد الإجمالي للوثائق الصالحة في نظام ما.

وفي امتداد لهذه الطريقة، غالبا ما يتم حساب متوسط الاستدعاء أو متوسط التحقيق، بالنسبة للوثائق المسترجعة حول موضوع ما، المتربعة على القمة س (س= ١٠، ١٥، ٢٠، ٣٠،...)، إذا كانت النتائج مرتبة طبقيا وفقا للصلاحية المنتظرة. وقد طبقت سلسلة مؤتمرات استرجاع النصوص **TREC** هذا المقياس على نطاق واسع، أي متوسط الاستدعاء أو متوسط التحقيق، عند الحدود القصوى للمخرجات **(Hartman, 1995)**. وقد قامت تشو وروزنتال **Chu and Rosenthal (1996)** بحساب متوسط التحقيق للنتائج العشر المسترجعة المتربعة على القمة، عندما كانتا بصدد تقييم نظم استرجاع الإنترنت آلتافيستا **AltaVista**، وإكزايت **Excite**، وليكوس **Lycos**. ويكفل متوسط الاستدعاء أو متوسط التحقيق، عند الحدود القصوى للمخرجات، بديلا عمليا لمقياسي الاستدعاء والتحقيق العاديين، وخصوصا عندما تكون مخرجات الاسترجاع مرتبة طبقيا، وكبيرة العدد.

والاستدعاء المعياري أو التحقيق المعياري **normalized**، شكل آخر للاستدعاء والتحقيق، يضع في الحسبان التتابع الذي تقدم به الوثائق المسترجعة للمستفيد؛ فمن الممكن مثاليا لنظام استرجاع المعلومات، أن يقدم جميع الوثائق الصالحة، قبل عرض أي من الوثائق غير الصالحة. وإذا ما أعطيت القيمة (١)لنظام مثالي، والقيمة (صفر) لأسوأ نظام ممكن، ذلك النظام الذي يمكن فيه تقديم جميع الوثائق غير الصالحة للمستفيد، قبل تقديم أي وثيقة صالحة، فإن تلك القيمة تصبح المعامل في حساب الاستدعاء أو التحقيق المعياريين. وقد وصف كل من كورفيج **Korfhage (1997)**، وفوسكت **Foskett (1996)** طريقتين مختلفتين لحساب الاستدعاء المعياري. كذلك يمكن حساب التحقيق المعياري على النحو نفسه. والافتراض الآخر بالنسبة لهذا المقياس، هو أن يكون المرء على دراية بكل الوثائق الصالحة في نظام ما، وهو مأزق مماثل بالنسبة لحساب الاستدعاء، ومن ثم فإن احتمال

هذا المقياس يعاني أيضا من ذلك الافتراض غير المحتمل. وفضلا عن ذلك، فإن الاستدعاء أو التحقيق المعياري، يتطلب في جوهره ترتيب مخرجات الاسترجاع طبقيا وفقا للصلاحية المنتظرة.

والاستدعاء النسبي هو نسبة الوثائق الصالحة المسترجعة التي فحصها المستفيد، إلى عدد الوثائق الصالحة التي يعلم المستفيد بوجودها في النظام **(Harter & Hert, 1997)**، أو يمكن أن يود فحصها **(Korfhage, 1997)**. لنفترض على سبيل المثال، أن المستفيد يعتقد أن هناك في النظام ٢٠ وثيقة صالحة حول استفسار معيّن، وبعد إجراء عملية البحث يسترجع المستفيد عشر وثائق، حكم عليها بأنها صالحة. حينئذ يكون الاستدعاء النسبي لعملية البحث هذه ٥٠ بالمئة. ومن الواضح أن الاستدعاء النسبي قد اقترح بقصد التغلب على المشكلات المرتبطة بالاستدعاء، إلا أنه هنا أيضا، يظل التساؤل قائماً، كيف يمكن للمستفيد أن يعرف، على سبيل المثال، عدد الوثائق الصالحة في نظام ما؟ والإجابة هي ما اقترح في ١١/ ١/ ٢/ ١/ ٢ بالنسبة لتحديد إجمالي عدد الوثائق الصالحة في النظام (كطريقة التجميع على سبيل المثال). وتعاود المشكلة القديمة نفسها الظهور بعد التحول إلى ما يسمى الاستدعاء النسبي.

ومقياس إي **E-measure**، الذي اقترحه سويتس **(1969) Swets** جمع موزون بين التحقيق والاستدعاء، لقياس فاعلية الاسترجاع، ومن الممكن حسابه بالمعادلة التالية:

$$E = 1-1 / [\beta P -1 + (1 + \beta) R - 1]$$

ويستعمل المعامل **β** في هذه المعادلة للتعبير عن التركيز على أي من التحقيق أو الاستدعاء؛ فعلي سبيل المثال فإن **β = 1** يقابل إضفاء أهمية متساوية على كل من التحقيق والاستدعاء. و**β = 0.5 and 2** تقابل إضفاء نصف أو ضعف الأهمية على الاستدعاء كما على التحقيق، على التوالي. وكلما انخفض مقياس إي **E** ارتفع مستوى

فاعلية نظام الاسترجاع. وقد طبق مقياس إي، على سبيل المثال، من جانب جريفيث ولاكهرست وويلت Griffiths, Luckhurst, and Willett (1986)، في دراستهم. ويمكن إبداء ملاحظتين بشأن مقياس إي؛ الأولى أنه يمكن أن يضفي درجات مختلفة من التركيز على أي من الاستدعاء أو التحقيق. والثانية، أنه يضم كلا من الاستدعاء والتحقيق في مقياس واحد للفاعلية. وكما سيتبين لنا فيما بعد، فإن مقياسا واحدا للتقييم هو المفضل كثيرا في أوساط استرجاع المعلومات.

وتكفل الأشكال المختلفة للاستدعاء والتحقيق، وجهات نظر وبدائل مختلفة لمقياسي التقييم. إلا أن أيا من الأشكال المختلفة لا يمكن أن يتغلب على نحو جوهري، وناهيك عن كلي، على المشكلات التي تواجه التحقيق والاستدعاء. ومن ثم ظهرت مقاييس تقييم إضافية (كالتسرب **fallout** على المثال) على مر السنين.

التسرب:

يعرف التسرب **fallout**، الذي طرحه سويتس **Swets (1963)** في البداية، بنسبة الوثائق غير الصالحة المسترجعة إلى جميع الوثائق غير الصالحة في مرصد بيانات النظام. ويتم تحديد التسرب (F) على النحو التالي باستعمال الرموز الواردة في الجدول رقم ١١/ ١:

$$F = \text{ب} / (\text{ب} + \text{د})$$

ويقيس التسرب عجز نظام استرجاع المعلومات عن استبعاد الوثائق غير الصالحة من نتائج الاسترجاع، الأمر الذي يسميه روبرتصون **Robertson (1969)** معامل الشوشرة **noise factor**. وتتراوح قيمة التسرب ما بين الصفر والواحد، شأنها في ذلك شأن الاستدعاء والتحقيق، إلا أنه كلما انخفضت قيمة التسرب، كان نظام استرجاع المعلومات أفضل من وجهة نظر التقييم.

وكما في حالة حساب الاستدعاء، فإن إجمالي عدد الوثائق غير الصالحة في النظام غير معروف، ومن ثم فإن ذلك العدد ينبغي تقديره، باتباع الطرق التي سبق أن نوقشت بالنسبة لتقدير الاستدعاء. ولا يبدو التسرب بوجه عام مقياس تقييم يستخدم بكثافة، على الرغم من أن روبرتصون **Robertson (1969)** صرح بأن بياني علاقة الاستدعاء بالتحقيق، (بياني يوقع قيم هذين المقياسين) لا يمكن تفسيره بسهولة كبياني علاقة الاستدعاء بالتسرب، وذهب إلى أن التحقيق ليس بالمقياس المفيد للأداء بنفس أهمية التسرب مرتبطاً بالاستدعاء.

العمومية:

تعرف العمومية **generality** بوصفها نسبة الوثائق في مرصد بيانات النظام، الصالحة لموضوع معين. وبناء على الرموز الواردة في الجدول رقم ١١/ ١، فإن ع (العمومية) يمكن حسابها على النحو التالي:

ع = (أ + جـ) / (أ + ب + جـ + د)

وكلما ارتفع مدى العمومية (أو كلما ازدادت كثافة الوثائق الصالحة في مرصد البيانات) ازداد ميل عملية البحث للسهولة **(Lancaster & Warner, 1993)**، إلا أن العمومية، على وجه التحديد، مقياس لجودة مرصد البيانات من منظور الصلاحية، وليست مقياسا لأداء الاسترجاع مباشرة. وفضلا عن ذلك، فإنه من الصعب دائما تحديد عدد الوثائق الصالحة التي لم تسترجع من النظام، أي قيمة جـ في الجدول رقم ١١/ ١.

المقاييس الأحادية لتقييم استرجاع المعلومات:

من بين انتقادات الاستدعاء والتحقيق عدم اكتمالهما كمقياسين للتقييم (راجع ١١/ ١/ ٢/ ١/ ٣)، وهذا هو السبب وراء استعمال الاستدعاء والتحقيق عادة،

بالجمع مع- وربما إلى جانب- معايير أخرى لأغراض التقييم. ويمكن للأمر أن يكون مثاليا، إن كان من الممكن استخدام مقياس تقييم واحد، بإدخال العوامل المتصلة في الحساب. ويكاد مقياس إي E يكون مثالا للمقياس الواحد لتقييم استرجاع المعلومات. فقد اقترح كوبر (Cooper (1973a, 1973b مقياسا للانتفاع utility، يقوم على الحكم الذاتي للمستفيد، على نظام استرجاع المعلومات، من حيث فائدته. ومن الممكن قياس الانتفاع من النظام بطريقة مختلفة، مادام ذلك يتفق ونظرية كوبر للمنفعة؛ فقد طلبت سو (Su (1991، على سبيل المثال، من المستفيدين تقييم الانتفاع بنتائج عمليات البحث، بناء على نوع ما من مقياس ليكرت Likert المكون من سبع درجات، وذلك بالنظر في قيمة نتائج عمليات البحث ككل. وقد تبين لها أن مقياس الانتفاع هذا، كان أفضل مقياس قائم بذاته للأداء التفاعلي لاسترجاع المعلومات، من بين عشرين مقياسا وقع عليها اختيارها من أجل دراستها.

كذلك يمكن أيضا تعريف مقاييس الانتفاع بطريقة مختلفة؛ كقيمة الوثائق المسترجعة ككل، في مقابل الوقت المستنفد، وتحدد القيمة بالدولارات، والإسهام في حل المشكلات (Saracevic, et al., 1988; Saracevic & Kantor, 1988). إلا أنه أيا كانت الطريقة التي يتم بها تعريف الانتفاع في التقييم، فإن هذا المقياس يعتمد دائما على الحكم الذاتي للمستفيد. وتفضي ذاتية هذا المقياس إلى نفس ما يبديه المهتمون من تخوف عند إصدار أحكام الصلاحية، لأجل حساب الاستدعاء والتحقيق.

مقاييس التقييم الأخرى لاسترجاع المعلومات:

عقب جريفيث (Griffith (1986 يوما، قائلا بأنه من الممكن استخدام ثلاثة أرقام فقط في تقييم نظم استرجاع المعلومات؛ عدد الوثائق الصالحة المسترجعة، وعدد الوثائق غير الصالحة المسترجعة، وإجمالي عدد الوثائق في مرصد بيانات النظام. ومن الممكن الحصول على هذه الأرقام جميعا دون أي تقدير أو تخمين. ويمكن لنسبة

الوثائق الصالحة المسترجعة إلى إجمالي عدد الوثائق في مرصد بيانات نظام ما، أن توضح قدرة النظام على العثور على الوثائق الصالحة. وكذلك الحال أيضا، يمكن لنسبة عدد الوثائق غير الصالحة المسترجعة إلى إجمالي عدد الوثائق في مرصد بيانات نظام ما، أن تبرز القدرة التمييزية للنظام؛ أي القدرة على رفض الوثائق غير الصالحة، إلا أن قليلا من مشروعات التقييم وضعت فكرة جريفيث في حيز التطبيق.

وقد مرت مقاييس التقييم بالنسبة لاسترجاع المعلومات برحلة طويلة ما تزال مستمرة. ومن بين جميع المقاييس التي نوقشت في هذا السياق، يحظى الاستدعاء والتحقيق بمعظم الاهتمام، فضلا عن معظم النقد أيضا. إلا أن هذين المقياسين هما المقياسان بحكم الواقع بالنسبة لمعظم جهود التقييم؛ نظرا لأنه من السهل تفسيرهما من حيث استرجاع الوثائق المرغوب فيها، ورفض الوثائق الغريبة (**Salton, 1992**). وسوف يظل الاستدعاء والتحقيق هما المقياسان الرئيسان للتقييم، بالنسبة لاسترجاع المعلومات، ما لم تتطور بدائل أفضل منهما وتثبت جدارتها.

معايير تقييم نظم استرجاع المعلومات:

الاستدعاء والتحقيق وغيرهما من المقاييس الأخرى التي سبقت مناقشتها، إنما هي لتقييم أداء الاسترجاع في الأساس. وأداء الاسترجاع كما يبدو، جانب واحد فقط من جوانب استرجاع المعلومات، وذلك على الرغم من أنه كما سبق أن تبين لنا، يوليه المستفيد النهائي أقصى اهتمام. إلا أنه لتكوين نظرة كاملة لمدى كفاءة أداء نظام استرجاع المعلومات، ينبغي وضع مقاييس أخرى خلاف تلك الخاصة بتقييم أداء استرجاع المعلومات في الحسبان. وينبغي استخدام مقاييس مختلفة لمختلف أنواع نظم استرجاع المعلومات. وتنظر الأقسام التالية في معايير التقييم بالنسبة لكل نوع من الأنواع الأربعة لنظم استرجاع المعلومات، وهي نظم الخط المباشر، ونظم الفهارس

المتاحة على الخط المباشر (OPACs)، ونظم استرجاع الإنترنت، ونظم استرجاع الوسائط المتعددة (وإن كانت نظم استرجاع الوسائط المتعددة لم تعالج كنوع متميز في الفصول السابقة).

معايير تقييم نظم الخط المباشر:

لقد كان لنظم الخط المباشر أطول تاريخ، بين الأنواع الأربعة الرئيسة لنظم استرجاع المعلومات (راجع الفصل الثامن)، ومن ثم فإنها كانت موضوع كثير من دراسات التقييم. كما أن معايير تقييم نظم الخط المباشر مصممة بطرق مختلفة.

ووفقا لما ذهب إليه لانكستر وورنر **Lancaster and Warner (1993)**، فإنه من الممكن لعمليات تقييم نظم الخط المباشر، أن تتم على ثلاثة مستويات: (١) عمليات تقييم الفاعلية، (٢) عمليات تقييم فاعلية التكلفة، (٣) عمليات تقييم عائد التكلفة. ويضع المستوى الأول، عمليات تقييم الفاعلية **effectiveness**، في الحسبان كلا من التكلفة، والوقت، وجودة النظام. وتحاول عمليات التقييم هذه الكشف عما إذا كان النظام يعمل بفاعلية فيما يتعلق بحاجة المستفيدين. ويربط المستوى الثاني، عمليات تقييم فاعلية التكلفة **cost-effectiveness**، عامل التكلفة بفاعلية النظام، بحيث يمكن المقارنة بين نظامين مختلفين لاسترجاع المعلومات بسهولة، في هذا المستوى. فكم على سبيل المثال، تبلغ تكلفة استرجاع وثيقة صالحة من نظام ما؟ ويمكن للضرب نفسه من المقارنة أن يكون من الصعب إجراؤه في المستوى الأول، نظرا لأن تلك المقاييس غير معيارية. أما المستوى الثالث، تقييم عائد التكلفة **cost-benefit**، فيذهب إلى ما هو أبعد من المستويين السابقين، وذلك بتقييم تكلفة تشغيل نظام استرجاع المعلومات، مقابل العائدات التي تتحقق نتيجة لوجود النظام. وكلما ارتفع مستوى التقييم، ازداد تعقد معايير التقييم. وتميل معايير المستوى الأول (أي زمن الاستجابة أو تغطية مرصد البيانات على سبيل المثال) لأن تكون مباشرة أو

صريحة. وتحتاج مقاييس المستوى الثاني إلى بعض الحسابات. وفي مقابل ذلك تتطلب معايير المستوى الثالث، إصدار الأحكام وتوافر البراهين، على نحو أكثر بكثير من التقييم البسيط. ومن الممكن تطبيق معايير التقييم، في جميع المستويات الثلاثة، في مشروع تقييم واحد، غير أنه في التطبيق العملي، تستخدم مقاييس المستويين الأول والثاني، ومقاييس المستوى الأول على وجه الخصوص، في معظم الأحيان، في تقييم نظم الخط المباشر؛ نظراً لجدوى هذه المقاييس وقابليتها للتطبيق.

وعلى الرغم من وضع معايير مختلفة لتقييم نظم الخط المباشر، يبدو أنه قد تم التوصل إلى إجماع حول أي الجوانب ينبغي أن تخضع للتمحيص. وقد لخص لانكستر وفاين **Lancaster and Fayen (1973)** هذه الجوانب في شكل ستة معايير: (١) التغطية، (٢) الاستدعاء، (٣) التحقيق، (٤) زمن الاستجابة، (٥) جهد المستفيد، (٦) شكل المخرجات. ومن عجب، أن نرى أن القدرة على إجراء عمليات البحث لم تدخل ضمن هذه المجموعة من معايير التقييم.

والتغطية معيار تقليدي لتقييم مرصد بيانات النظام. وتشمل الأوجه المحددة لتغطية مرصد البيانات أنواع الوثائق (كبراءات الاختراع، ومقالات الدوريات، وتقارير البحوث على سبيل المثال)، وعدد الوثائق، وتتابع التجديد، والمدى الزمني للتغطية الراجعة. وتحدد تغطية مرصد بيانات النظام ما يمكن استرجاعه فيما بعد من النظام.

والاستدعاء والتحقيق جزء من معايير التقييم لقياس أداء الاسترجاع. وقد أصبح هذان المقياسان، كما بينا في ١١/ ١/ ٢ المكونات التقليدية في معظم جهود التقييم، إلا أنه لم يتم وضع مواصفة معيارية لمعدلات الاستدعاء والتحقيق المرضية، وذلك لسببين؛ أولهما أنه من الممكن للمستفيدين المختلفين أن تكون لهم توقعات مختلفة لكل من الاستدعاء والتحقيق، في الظروف المختلفة؛ فإذا كان هناك على سبيل

المثال مستفيد ينوي إعداد مراجعة علمية للوضع الراهن للمعرفة في موضوع ما، فإن الاستدعاء المرتفع يمكن أن يكون هو المتوقع. وإذا كان المستفيد لا يهتم إلا بأحدث التطورات في مجال ما، فإنه يمكن للتحقيق المرتفع أن يكون هو المتوقع. والسبب الآخر لغياب مواصفة معيارية لمعدل الاستدعاء والتحقيق الذي يمكن قبوله، هو علاقة التناسب العكسي التي يفترض وجودها بين المقياسين؛ فممن الممكن للزيادة في أحدهما أن تفضي إلى تناقص في الآخر. ومن ثم فإنه من الصعب إيجاد حد أمثل لكل من الاستدعاء والتحقيق. يضاف إلى ذلك أن لحجم مرصد بيانات النظام تأثيره على وضع حد للاستدعاء والتحقيق المرضيين. ومن الممكن لمعدلات التحقيق المنخفض أن تكون أقل قابلية للتغاضي عنها، في مراصد البيانات الأكبر، مما هي عليه في مراصد البيانات الأصغر. فمن الممكن على سبيل المثال لمستفيد ما أن يكون على استعداد للنظر في ستين تسجيلة مسترجعة، للتحقق من أن بينها عشرون صالحة، إلا أنه قد يكون أقل استعداداً بكثير، لأن ينظر في ستمئة تسجيلة للعثور على مئتين، على الرغم من أن معدل التحقيق في كلا الحالتين، قد يكون واحداً (٣٣ بالمئة) **(Lancaster & Waner, 1993)**.

ويدل زمن الاستجابة على الوقت المستنفد بين تقديم الاستفسار وتلقي النتائج. ويبدو هذا المعيار لا غنى عنه، وخصوصا في نظم الخط المباشر في أيامها المبكرة، عندما كانت تقنيات المعلومات (كالاتصالات بعيدة المدى، وطاقة الحاسبات على سبيل المثال) في مرحلة طفولتها، وكان استرجاع المعلومات، اعتمادا على نظم الخط المباشر باهظ التكلفة. فإذا كان بإمكان نظام استرجاع المعلومات تجهيز الاستفسارات بسرعة، فإنه يمكن الحد من تكلفة استخدامه، الأمر الذي يفسر أهمية وجود هذا المعيار في التقييم.

وينطوي جهد المستفيد كمعيار للتقييم، على عدة عناصر، جميعها يقيس مقدار الجهد الذي يتعين على المستفيد أن يبذله، لإجراء عمليات البحث في بيئة الاسترجاع على الخط المباشر؛ فكم من الوقت، على سبيل المثال، يستغرقه إلمام المستفيد بالنظام أو التآلف معه؟ ما مدى تعاطف تفاعل المستفيد مع النظام؟ أي أنواع الدعم (كالموجزات الإرشادية المطبوعة، أو المساعدة على الخط المباشر على سبيل المثال) تتوافر للمستفيد؟ وكلما قل الجهد الذي يحتاج المستفيد إلى بذله، كان النظام أفضل. ومن الخواص الفريدة لنظم الخط المباشر، أن اختصاصيي المعلومات كوسطاء، غالبا ما يجرون عمليات البحث نيابة عن المستفيدين النهائيين، كما أنهم مدربون على نحو أفضل من تدريب هذه الفئة الأخيرة. وينبغي تذكر هذا الاختلاف، عندما يكون المرء بصدد النظر في عامل جهد المستفيد، في التقييم.

ويمثل شكل المخرجات المعيار الأخير، الذي وضع للتقييم في نظم الخط المباشر؛ فمن الممكن تقديم مخرجات عمليات البحث على الخط المباشر بطرق مختلفة (كالمعلومات الوراقية فقط) أو المعلومات الوراقية مصحوبة بمستخلصات، على سبيل المثال). وعادة ما تكفل نظم الخط المباشر الكثير من خيارات المخرجات، إلا أنها لا ترتب تلقائيا نتائج البحث طبقيا؛ نظرا لأن استخدام مقومات الترتيب الطبقي (كالأمر **RANK** في ديالوج **DIALOG**، على سبيل المثال) عادة ما يكون باهظ التكلفة. وعادة ما تقدم نتائج الاسترجاع من نظم الخط المباشر، في ترتيب زمني تنازلي؛ أي أن تأتي أحدث تسجيلات أضيفت إلى مرصد البيانات في الصدارة. وكما تبين لنا في ١٠/ ٣/ ٢/ ٢، فإن المرونة هي الكلمة المفتاح للحصول على أفضل تقييم في هذا الصدد.

وقصارى القول، فإن التغطية، والاستدعاء، والتحقيق، وزمن الاستجابة، وجهد المستفيد، وشكل المخرجات، قد استخدمت كمعايير تقييم لنظم الخط المباشر،

على الرغم من وجود بعض مظاهر الاختلاف أحيانا. ويمكن لهذه المعايير أن تكفل أيضا أساسا لتقييم الأنواع الأخرى لنظم استرجاع المعلومات.

معايير تقييم الفهارس المتاحة على الخط المباشر:

تمثل الفهارس المتاحة على الخط المباشر **OPACs**، المصممة للعثور على المعلومات الوراقية المتعلقة بمجموعة مكتبة ما، نوعا فريدا من نظم استرجاع المعلومات. وقد تطورت معايير التقييم الخاصة بهذه الفئة من النظم، في المقام الأول، عن طريق الدراسات المقارنة، ويبدو التقرير الذي أعده هايلدرث **Hildreth (1982)**، هو الأقوى أثرا في هذا التطور. بل إن جونز **Jones** (في **Hildreth, 1982, p.xii**) قد تكهن بأن أي جهد تالٍ في هذا المجال،لم يكن بمنأى عن التأثر بتقرير هايلدرث. وقد شكل الإطار الذي قدمه في هذا التقرير، فعلا الجهود التي تلت، في تقييم الفهارس المتاحة على الخط المباشر (راجع على سبيل المثال **O'Rourke, 1987; Salmon, 1983**) وذلك على الرغم من أن البحث لم يقصد به أن يكون تقييما لعشرة فهارس متاحة على الخط المباشر، تمت دراستها (**Hildreth, 1985**). وبعد استشارة البحوث الصالحة في الموضوع (مثل **Hildreth, 1982**)، أعد أورورك **O'Rourke (1987)**، لأغراض التقييم، قائمة مراجعة، من ٩٥ سؤالا، مقسمة على خمس فئات هي: (١) الخواص التشغيلية، (٢) نقاط الوصول، (٣) خواص البحث، (٤) خواص العرض، (٥) خواص مساعدة المستفيد. ومن الممكن لهذه الفئات الخمس أن تشكل على نحو معقول، معايير تقييم للفهارس المتاحة على الخط المباشر.

ويتعلق معيار الخواص التشغيلية بكيفية تفاعل المستفيد فعلا مع الفهرس المتاح على الخط المباشر؛ كيف، على سبيل المثال، يصل المستفيد إلى النظام؟ هل يكفل النظام للمستفيد اختيار هدف (كمجموعة المكتبة المحلية، أو مجموعة تكتل للمكتبات، على سبيل المثال) لإجراء عمليات البحث، بالإضافة إلى الخيار الافتراضي **default**؟ هل

بإمكـان المـستفيد تعـديل البحـث أو تغيـير أوضـاع النظـام الافتراضـية (كعـرض التسجيلات الموجزة على سبيل المثال)؟ هل من السهل بالنسبة للمستفيد الانتقال من مستوى عرض إلى آخر، أو من إحدى التسجيلات إلى ما تليها أو مـا سـبقتها؟ ولقـد كانت واجهة التعامل **interface** (أي لغة إصدار الأوامر في مقابل الاختيار من قائمة الخيارات) من الموضوعات الجوهرية عندما وضعت معايير التقييم تلك، إلا أنها تبدو أقل صلاحية بالنسبة للفهارس المتاحة على الخط المباشر في الوقت الـراهن؛ نظـرا لأن قليلا من هذه النظم لا يـزال يطبـق لغـة إصـدار الأوامـر. ولقـد كـان للتحـول مـن الفهارس البطاقية إلى الفهارس المتاحة على الخط المباشر، أثره البالغ في تفاعل المستفيد مع النظام، الأمر الذي يبدو مبررا رئيسا لاختيار هذا المعيار لتقييم الفهـارس المتاحـة على الخط المباشر.

وتحدد نقاط الوصول كيف يمكن البحث في معلومات الفهارس المتاحـة عـلى الخط المباشر، وتشكل معيارا آخر لتقييم هذا الشكل مـن الفهـارس؛ ففـي الـسنوات المبكرة لتطور هـذه الفهـارس (أي ثمانينيـات القـرن العـشرين) كـان اسـم المؤلـف، والعنـوان، ورقـم الاسـتدعاء **call number**، والـرقم المعيـاري الـدولي للكتـاب **International Standard Book Number (ISBN)**، والرقم المعيـاري الـدولي للدوريـة **International Standard Serial Number (ISSN)**، هـي نقـاط الوصول المألوفة. ثم أضيفت الكلمات المفتاحيـة، ورؤوس الموضـوعات بعـد ذلـك. ومن الناحية النظرية، فإنه كلما زاد عدد نقاط الوصول التي يكفلها الفهرس المتاح على الخط المباشر، كان من الأيسر على المستفيد العثـور عـلى المعلومـات في النظـام، إلا أن بعض نقاط الوصول (كالرقم المعياري الدولي للكتاب **ISBN**، ورؤوس الموضوعات على سبيل المثال) لا تستخدم بكثافة، لأن المستفيدين، وخـصوصا المبتـدئين مـنهم، لا يعرفون الكثير عن مضمون هذه النقاط. وفي مقابل ذلك، فإن نقاط الوصول في حـد

ذاتها، لم تصبح معيارا لتقييم نظم الخط المباشر؛ نظرا لأن هناك الكثير من هذه النقاط في بيئة الاسترجاع هذه. ومن الممكن اختيار البحث بالحقل، ضمن مقومات البحث، بدلا منها، لتغطية البعد الخاص بنقاط الوصول.

وتختلف خواص البحث في الفهارس المتاحة على الخط المباشر، من نظام لآخر. وعندما يتم تقييم الفهارس المتاحة على الخط المباشر، من منظور خواص البحث، تتم مراعاة عناصر متعددة؛ أولها أي أنواع مقومات البحث تحظى بالدعم؟ هل جميع المعاملات البولينية الثلاثة متاحة؟ ماذا عن البحث بالبتر، والبحث بالتقارب؟ ثانيا، هل يكفل النظام أي خواص تتصل بالبحث، كقائمة رؤوس الموضوعات، على الخط المباشر، والإحالات، ومقترحات الارتفاع بمستوى عمليات البحث، على سبيل المثال؟ ثالثا، هل يمكن للمستفيدين تحديد عمليات بحثهم بسنة النشر، أو نوع الوثيقة، أو ما شابه ذلك؟ وكلما زادت الخيارات التي يمكن للنظام دعمها كان النظام أفضل.

ويتم اختيار خواص العرض، وهي مفهوم أعرض من خيارات المخرجات، كمقياس لتقييم الفهارس المتاحة على الخط المباشر. ومن الممكن تحليل هذا المعيار إلى طول التسجيلة (أي العرض الموجز أو الكامل على سبيل المثال)، وعدد التسجيلات في كل قائمة، والخواص الأخرى المتصلة بالعرض. كذلك يمكن عرض التسجيلات كاملة بصيغ مارك **MARC**، وهذه خاصية تستخدم في معظم الأحيان، من جانب اختصاصيي المعلومات. وقد طرح أورورك **O'Rourke (1987)** المزيد من الأسئلة المحددة لاختبار جودة عرض الفهرس المتاح على الخط المباشر. فهل يمكن على سبيل المثال، للنظام أن يعرض تلقائيا التسجيلة عندما تكون هي الإجابة الوحيدة لمصطلح البحث؟ هل التيجان (مثل **TI** للعنوان، على سبيل المثال) واضحة في حد ذاتها، في حالات العرض المميزة للتسجيلات؟ هل تخلو حالات عرض التسجيلات من

المختصرات، والأكواد، والمصطلحات الخاصة؟ وفضلا عن ذلك، يمكن توجيه أسئلة حول نوع الحرف وأسلوبه وحجمه أو أشكال الحروف وأحجامها المستخدمة، في حالات عرض نتائج البحث في الفهرس المتاح على الخط المباشر. فلا يوصي على سبيل المثال بكل أحجام الحروف الكبيرة، وأنواع الحروف الصغيرة (كبنط ١٠ مثلا) على سبيل المثال، لعرض البيانات في الفهارس المتاحة على الخط المباشر.

وتشكل مساعدة المستفيد معيارا آخر لتقييم الفهارس المتاحة على الخط المباشر. وكلما ازدادت المساعدة التي يمكن للنظام تقديمها للمستفيد، تراجعت احتمالات شعور المستفيد بأنه لا حول له ولا طول، عندما يتفاعل مع الفهرس المتاح على الخط المباشر. ومن الممكن تصميم مقومات كثيرة لمساعدة المستفيدين من هذا الشكل من الفهارس. وتشمل الأمثلة رسالة تمهيدية تصف الفهرس المتاح على الخط المباشر بوجه عام، وقائمة بالملفات التي يمكن الوصول إليها أو الحقول القابلة للبحث، وتفسيرات رسائل الخطأ، والتوعية على الخط المباشر، والمساعدة المتصلة بعملية بحث بعينها على الخط المباشر. كذلك ينبغي مراجعة الفهرس المتاح على الخط المباشر أيضا، لمعرفة ما إذا كان بإمكانه بيان الخلفية التاريخية للبحث، على سبيل المثال، وما إذا كان يقدم تعقيبات إجرائية أو إرشادية بشكل روتيني. ولا ينبغي وضع أي افتراضات حول تآلف المستفيد مع النظام، عند مناقشة مساعدة المستفيد. وينبغي أن تكون رسائل المساعدة واضحة وبسيطة ودقيقة.

وقد تطورت معايير التقييم الخمسة هذه في ثمانينيات القرن العشرين، عندما بدأ استخدام الفهارس المتاحة على الخط المباشر في المكتبات. وقد تم إدخال تحسينات وخواص جديدة، ينبغي أن توضع في الحسبان في تقييم الطبعات الحديثة من الفهارس المتاحة على الخط المباشر، وخصوصا الجيل الناشئ. فمعظم الفهارس المتاحة على الخط المباشر متشابكة الآن، على سبيل المثال، بحيث تكفل التعامل عن بعد، من المنزل أو

المكتب أو أي أماكن أخرى خلاف المكتبة، فهل التعامل عن بعد ملائم؟ هل تتوافر أشكال دعم ملائمة لمساعدة المستفيد في التعامل عن بعد؟ ومن ناحية أخرى، فإنه كما بينا في ٨/ ٣، يقوم عدد كبير من الفهارس المتاحة على الخط المباشر الآن، مقام بوابات لمصادر معلومات ونظم أخرى لاسترجاع المعلومات. فما المقومات المتاحة للمستفيدين للإفادة من هذه الخاصية؟ وينبغي أن تحظى هذه الخواص والتطورات الجديدة للفهارس المتاحة على الخط المباشر بالاهتمام لأغراض التقييم، بالإضافة إلى المعايير الخمسة التي تناولناها في هذا السياق.

ولقد اكتسب ما يسمى بالجيل التالي من الفهارس المتاحة على الخط المباشر فعلا، خواص جديدة لم تتوافر في الطبعات المبكرة من هذه الفهارس. وتشمل نماذج مثل هذه الخواص التلقيم المرتد الخاص بالصلاحية، وترتيب نتائج عمليات البحث طبقيا، والبيئة المعدلة للتصفح، ومراجعة هجاء استفسارات البحث **(Antelman, Lynema & Pace, 2006; Markey, 2007)**. ويتم إجراء تقييم الجيل التالي من الفهارس المتاحة على الخط المباشر، الآن في معظم الأحيان، عند تقييم الفهرس المتاح على الخط المباشر حديث النشأة (راجع على سبيل المثال **Antelman, Lynema & Pace, 2006; Lombardo & Condic, 2000**)، ومع ذلك فإنه لم يوضع بعد إطار جديد لهذا الفرص. إلا أن هذا الموقف سوف يتغير ولاشك، نتيجة لتطبيق المزيد من خواص الجيل التالي من الفهارس المتاحة على الخط المباشر، وخصوصا عندما أعرب كثيرون (مثل، **Marcum, 2006; Markey, 2007**) فعلا عن مخاوفهم بشأن مصير هذا النوع من الفهارس في ظل جوجل العملاق.

معايير تقييم نظم استرجاع الإنترنت:

تناول عدد قليل نسبيا من مشروعات التقييم، نظم استرجاع الإنترنت، أحدث الأعضاء في أسرة نظم استرجاع المعلومات، على الرغم من أن بعض الباحثين (مثل

Chu & Rosenthal, 1996; Ding & Marchionini, 1996; Leighton, 1995; Leighton & Srivastava, 1999) حاولوا اختبار أداء استرجاع هذه الفئة من النظم. ولما كان أداء الاسترجاع يمثل جانبا واحدا فقط لنظم استرجاع المعلومات، فإن الأبعاد الأخرى لنظم استرجاع الإنترنت ينبغي أن ؛ ع أيضا في الحسبان في التقييم.

وقد اقترحت تشو وروزنتال **Chu and Rosenthal (1996)** إطار تقييم لمحركات بحث العنكبوتية العالمية، النوع الغالب لنظم استرجاع الإنترنت، حتى وقت تأليف هذا الكتاب. وهذا الإطار مشتق من معايير تقييم سبعينيات القرن العشرين، التي وضعها لانكستر وفاين **Lancaster and Fayen (1973)** لنظم الخط المباشر، مع إضافات ومراجعات مصممة خصوصًا لنظم استرجاع الإنترنت. ونظرا للتغيرات التي طرأت على نظم استرجاع الإنترنت بوجه عام، تجاوز الزمن الآن بعض الجوانب المحددة في معايير التقييم الشاملة الخمسة (أي تكوين كشافات العنكبوتية، وقدرات البحث، وأداء الاسترجاع، والمخرجات، وجهد المستفيد) وقت إجراء الدراسة، إلا أن الإطار النظري لا يزال صالحا لتقييم نظم استرجاع الإنترنت.

وتكوين الكشافات في نظام استرجاع المعلومات، عنصر مهم للتقييم؛ نظرا لأن محتوى الكشافات يحدد كيف يمكن استرجاع ما يشتمل عليه النظام من معلومات. فكلما وجه طلب للبحث في العنكبوتية، على سبيل المثال، فإن كشاف العنكبوتية، وليس صفحات العنكبوتية نفسها، هو ما تتم مراجعته لاسترجاع المعلومات. والتغطية، وتتابع التجديد، وقطاعات وثائق الإنترنت المكشفة (كالعنوان بالإضافة إلى السطور الأولى، أو الوثيقة برمتها، على سبيل المثال) هي العناصر الثلاثة التي يتم اختبارها. وبينما يتوقف مدى المكونات الثلاثة إلى حد بعيد، على طاقة العتاد ومدى تطوره، والبرمجيات التي تدير الكشاف أو مرصد البيانات، فإن التغطية الضخمة،

والتجديدات المتتابعة، وتكشيف النصوص الكاملة، ليس من الضروري أن تسفر عن نظم لاسترجاع الإنترنت، أفضل من المقاييس الأخرى؛ فهناك عوامل أخرى (كواجهة التعامل ومقومات البحث عن سبيل المثال) تنهض بدور في جودة نظم استرجاع الإنترنت.

وتشكل مقومات البحث معيارا آخر لتقييم نظم استرجاع الإنترنت؛ نظرا لأن الاختلاف هو القاعدة وليس الاستثناء في هذه الحالة. ومقومات البحث الأساس والمألوفة قد لا تكون موجودة في معظم نظم استرجاع الإنترنت. ومن ناحية أخرى، فإن مقومات الاسترجاع (كالبحث المشوش على سبيل المثال) التي كان من الممكن أن نجدها فقط في المختبرات في الماضي، تطبق الآن في بعض نظم استرجاع الإنترنت. فنظام استرجاع الإنترنت الكفء ينبغي أن يتضمن مقومات البحث الأساس، التي ألفها المستفيدون، وتشمل المنطق البوليني، والبحث بالعبارات، وإمكانية التحديد (كالتحديد بالحقل على سبيل المثال). ولما كانت مقومات البحث في نظام استرجاع الإنترنت، هي التي تتحكم في أدائه في النهاية، فإن من شأن غياب هذه المهام الأساس أن تعوق النظام بشدة.

وعادة ما يتم تقييم أداء الاسترجاع بثلاثة معايير؛ هي التحقيق، والاستدعاء، وزمن الاستجابة. وبينما يمكن قياس هذه المتغيرات الثلاثة جميعا بطريقة كمية، فإنه ينبغي مراعاة المزيد من الحيطة والحذر، عندما يكون المرء بصدد الحكم على صلاحية المواد المسترجعة، وتقدير إجمالي عدد الوثائق الصالحة لموضوع بعينه، في نظام قائم على شبكة. ونظرا لأن تمثيلات أو بدائل وثائق الإنترنت (كالملخصات أو الاقتباسات على سبيل المثال) يتم إعدادها بوسائل آلية في غالب الأحيان، فإن أحكام الصلاحية القائمة على مثل هذه التمثيلات، يمكن أن تسبب المزيد من المخاوف بشأن إمكانية التعويل عليها. يضاف إلى ذلك أن الطابع الديناميكي المتقلب للإنترنت يجعل من

الصعب تقريبا، تحديد إجمالي عدد الوثائق الصالحة في نظام لاسترجاع الإنترنت. ومن ثم، فإن حساب الاستدعاء أصعب في بيئة الإنترنت، مما هو عليه في الأنواع الأخرى من نظم استرجاع المعلومات. إلا أن أوبنهايم ورفاقه **Oppenheim, et al. (2000)** يرون أن الاستدعاء النسبي يمكن حسابه باتباع الطريقة التي اقترحها كلارك وويليت **Clarke and Willett (1997)** عندما كانا بصدد تقييم أداء نظم استرجاع الإنترنت. وفيما يتعلق بزمن الاستجابة كأحد عوامل الأداء، فإنه في سبيله لأن تتراجع أهميته نظرا لأن نظم استرجاع الإنترنت تزداد تحسنا من حيث طاقة التجهيز.

وينبغي النظر في المخرجات كمعيار للتقييم بالنسبة لنظم استرجاع الإنترنت، وذلك من ثلاث زوايا، هي إمكانية الوصول أو التعامل، والمحتوى، والشكل. وبفضل خاصية الروابط الفائقة، يستطيع المستفيد الوصول إلى الوثائق الكاملة لناتج الاسترجاع، إذا كانت الروابط تعمل على النحو السليم. إلا أن هناك روابط مختلة، كالروابط المنقطعة (كالروابط بالمواقع التي تغيرت المحددات الموحدة لأماكن المصادر **URLs** الخاصة بها، على سبيل المثال)، أو الروابط الميتة (كالروابط بالمواقع التي لم تعد متاحة، على سبيل المثال). وينبغي للنظام المثالي لاسترجاع الإنترنت، أن يجعل عدد الروابط المختلة في حده الأدنى، وبذلك يكفل الحد الأقصى لإمكانية الوصول للنتائج النهائية. ويتحدد محتوى المخرجات، إلى حد ما، تبعاً للطريقة التي بني بها نظام استرجاع الإنترنت؛ أي ما إذا كان النظام يقتطف مجرد اقتطاف، أم يكشف فعلا ما يتضمنه عند بناء مرصد بياناته؟ وينبغي أن تكون الإجابة واضحة؛ فالمقتطفات لا تكفل مساعدة تذكر للمستفيد في تحديد "موضوع أو محتوى **aboutness**" النتائج المسترجعة. وهناك جانب آخر يتصل بمحتوى المخرجات، وهو مقدار التكرار في النتائج. ولا يستطيع سوى بعض ما وراء نظم استرجاع الإنترنت، في الوقت الراهن، التخلص من المكررات. إلا أن المخاوف بشأن أشكال مخرجات نظم استرجاع

الإنترنت، قلما تمثل قضية تذكر؛ نظرا لأن النظم تميل لأن تحدد سلفا جميع أشكال المخرجات المحتملة، ومن الممكن الاطلاع على الوثائق؛ الأصلية التي تمثلها المخرجات بسهولة إذا رغب المستفيد في ذلك، ومن ثم، فإن إمكانية الوصول والمحتوى هما خاصتا المخرجات الفريدتان والمهمتان، بالنسبة لتقييم نظم استرجاع الإنترنت.

ويدل جهد المستفيد في السياق الراهن، على التوثيق وواجهة التعامل؛ نظرا لأن هذين العاملين يؤثران على نحو جوهري في جهود المستفيدين، في الإلمام بنظم استرجاع الإنترنت، والإفادة منها؛ فالتوثيق المعد جيدا، وواجهة التعامل التي تتعاطف مع المستفيد، ينهضان بدور لا يستهان به في اختيارات المستفيدين لنظم استرجاع الإنترنت. وعلى عكس الأنواع الأخرى لنظم استرجاع المعلومات، فإن نظم استرجاع الإنترنت مجانية بوجه عام، ولا تبدو أوجه التمييز فيما بينها واضحة، ومن ثم فإن جاذبية أي من نظم استرجاع الإنترنت، يتم التعبير عنها للمستفيدين، عن طريق واجهة التعامل والتوثيق في المقام الأول. وبعبارة أخرى، فإن المستفيدين لن يقع اختيارهم على نظام لاسترجاع الإنترنت، ما لم يكونوا يشعرون بالراحة مع واجهة التعامل الخاصة به، وما لم يكونوا قادرين على قراءة واستيعاب توثيقه عندما يستشيرونه. وقد حاولت نظم استرجاع الإنترنت كفئة، الحد من جهود المستفيدين، بتطبيق أسلوب تعبئة الاستمارات، والطرق التصويرية للتفاعل، وبتوفير مقومات المساعدة على الخط المباشر. وعند تقييم نظم استرجاع الإنترنت من وجهة نظر جهد المستفيد، فإن المرء بحاجة لأن يختبر، على وجه الخصوص، ما إذا كانت الشاشة مزدحمة جدا، أو ما إذا كانت رسائل المساعدة مغرقة في التقنية أو الإيجاز؛ نظرا لأن مثل هذه المخاوف تبدو عادية بالنسبة لبعض النظم. ومن ناحية أخرى، يختلف المستفيدون من نظم استرجاع الإنترنت، عن أولئك الذين يفيدون من النظم الأخرى لاسترجاع

المعلومات. فهم أكثر الفئات تنوعا من حيث مهارات وخبرات استرجاع المعلومات. وقد لا يكونون قد تلقوا أي شكل من التدريب قبل التعامل مع نظم استرجاع الإنترنت فعلا. يضاف إلى ذلك أنه قلما يكون هناك عادة وسيط، كما في حالة نظم الخط المباشر، ليقوم بإجراء عملية البحث للمستفيد. ويفرض غياب الوسيط، وافتقار معظم المستفيدين من نظم استرجاع الإنترنت إلى الخبرة، شرطا بالنسبة للحد من جهد المستفيد، أكثر صرامة مما يتعلق بأي بيئة أخرى لاسترجاع المعلومات.

وفضلا عن معايير التقييم الخمسة التي سلف ذكرها، ينبغي أيضا وضع استرجاع العبارات، والاسترجاع متعدد اللغات، والبحث عبر اللغات، في الحسبان، في تقييم نظم استرجاع الإنترنت. وقد شرحنا مفهوم استرجاع المقاطع أو الفقرات في ١/٢/١، قبل مناقشة نظم استرجاع الإنترنت على وجه التحديد. ولما كانت العنكبوتية قد أصبحت أحد التطبيقات الجوهرية للإنترنت، وتتخذ وثائق العنكبوتية شكل الروابط الفائقة، فإنه يبدو من الضروري، على نحو متزايد، توافر مقومات استرجاع الفقرات أو المقاطع في الإنترنت؛ فالنهج التقليدي لاسترجاع الوثائق كاملة، لا يكفل مساعدة تذكر بالنسبة للمستفيد من نظم استرجاع الإنترنت، في تحديد المعلومات التي تدعو الحاجة إليها؛ ويرجع ذلك إلى حد ما إلى البنيان الفائق لوثائق العنكبوتية. فمن الممكن لمقطع أو فقرة المعلومات المرغوب فيه، أن يكون مطموراً في أعماق البنيان، كما أن الحجم غير العادي لبعض وثائق العنكبوتية، يعد مشكلة في حد ذاته. ولهذا السبب يبدو استرجاع الفقرات أو المقاطع خاصية مفضلة بالنسبة لنظم استرجاع الإنترنت. وفي عام ١٩٩٧، بدأت سلسلة مؤتمرات استرجاع النصوص **TREC** الشهيرة، مشروع بحث حول استرجاع الفقرات سمي مسار الرد على الأسئلة **Question Answering (Voorhess, 2000)**.

ويتعلق الاسترجاع متعدد اللغات، أو البحث عبر اللغات، الذي نوقش في ٩/ ١، بالبعد اللغوي لنظم استرجاع المعلومات. وكما عولجت في الفصل التاسع، فإن المعلومات باللغات غير الإنجليزية، يتم تحميلها، على نحو متزايد، على الإنترنت، بما يوحي بأن عامل اللغة في سبيله لأن يصبح متزايد الأهمية، بالنسبة لنظم استرجاع الإنترنت. وتتضمن أنواع أخرى من نظم استرجاع المعلومات، معلومات متعددة اللغات، إلا أنها لا تقترب إلى مستوى يمكن مقارنته بمستوى نظم استرجاع الإنترنت. وكان المستفيدون غير الناطقين بالإنجليزية، يمثلون أكثر من ٧٠ بالمئة من إجمالي المتعاملين مع الإنترنت، في يونيو عام ٢٠٠٨ **(Miniwatts Marketing Group, 2009)**، والعدد في ارتفاع مستمر. ويبدو من المعقول فقط إذن، إدخال البعد اللغوي في تقييم نظم استرجاع الإنترنت. وقد تمت دراسة كل من الاسترجاع متعدد اللغات، والبحث عبر اللغات في مؤتمرات استرجاع النصوص **TREC** **(Voorhess & Harman, 2000)**. وقد لخص كلاف **Clough (2007)** الجهود التي بذلت في منتدى كلف للصور **Image CLEF**، وهو مسار لمنتدى التقييم عبر اللغات **Cross-language evaluation forum (CLEF)**، نشأ للتقييم واسع المدى لنظم استرجاع الصور عبر اللغات. وقد ورد وصف منتدى التقييم عبر اللغات **CLEF** في ٩/ ١/ ٣.

ويتصل كل من استرجاع المقاطع أو الفقرات **passage retrieval**، والاسترجاع متعدد اللغات، والبحث عبر اللغات، في الواقع، بمعيار تقييم مقومات البحث. إلا أن هذه السبل قد نوقشت بمعزل بعضها عن بعض، في هذه الحالة؛ لتأكيد أهميتها، وإبراز التغيرات التي طرأت على إطار التقييم الذي اقترحته تشو وروزنتال **Chu and Rosenthal (1996)** لنظم استرجاع الإنترنت.

وفي ذروة أنشطة التقييم بالنسبة لنظم استرجاع الإنترنت، أشار أوبنهايم وآخرون (2000) .Oppenheim, et al، إلى وجود قليل من الاطراد فيها بين الدراسات المتصلة، من حيث معايير أو طرق التقييم المتبعة. وقد أوصوا بتطوير مجموعة معيارية من الأدوات اللازمة لعمليات التقييم في المستقبل، كما قدموا أيضا بعض المقترحات المبدئية حول ما ينبغي أن تتضمنه هذه المجموعة المعيارية من الأدوات. ولما كانت نظم استرجاع الإنترنت، هي أسرع الأنواع نموا في أسرة نظم استرجاع المعلومات، فقد آن الأوان لتجديد جهودنا البحثية في تقييم هذه النظم، لدعم مسيرة تقدمها ونموها.

معايير تقييم نظم الاسترجاع متعددة الوسائط:

ناقشنا استرجاع المعلومات متعددة الوسائط في ٩/ ٢، تحت العناوين الفرعية "استرجاع الصور الثابتة" و"استرجاع الأصوات"، و"استرجاع الصور المتحركة". وتشهد جهود البحث والتطوير في هذا المجال توسعا سريعا، مسايرة لنضج التقنيات المتصلة به. ونتيجة لذلك أصبح الآن هناك عدد لا بأس به من نظم الاسترجاع الخاصة بالمعلومات متعددة الوسائط، في متناول المستفيدين. وقد لوحظت الحاجة إلى تقييم مثل هذه النظم الخاصة بمعلومات الوسائط المتعددة، من جانب الباحثين في المجال. وفي التقرير الذي أعده ترانت (2004) **Trant**، تدل مسارات ترِك **TREC** الخاصة باسترجاع الوثائق الناطقة **TREC tracks SDR (Spoken Document Retrieval)**، والفيديو (التي سبقها عام ٢٠٠٣ تركفيد **TRECVid; www-nlpir.nist.gov/projects/trecvid**)، بالإضافة إلى مشروعات مثل إميجكلف **Image CLEF (imaglclef.org)**، تدل جميعها على ضرورة تقييم استرجاع المعلومات متعددة الوسائط.

وعلى عكس تقييم استرجاع المعلومات النصية، الذي تراكمت محاولاته على مدى خمسين عاما من الجهد والخبرة (مثل تجارب كرانفيلد **Cranfield** الرائدة) لا يبدو هناك سوى قليل من الباحثين الحريصين على محاولة تقييم استرجاع المعلومات متعددة الوسائط. ويفسر هذا العزوف، إلى حد ما، لما بدأ مجلس موارد المكتبات والمعلومات **Council on Library and Information resources**، وكذلك اتحاد المعلومات المتشابكة **Coalition for Networked Information**، وكلاهما في الولايات المتحدة الأمريكية، مشروعا يهدف لتطوير خدمات مرصد بيانات استرشادي لاسترجاع الصور (**Trant, 2004**). والتجارب الفردية لاختبار نظم بعينها لاسترجاع المعلومات متعددة الوسائط مألوفة (راجع على سبيل المثال **Chang, 2002; Deselaers, Keysers & Ney, 2008**)، في الوقت الذي غالبا ما كانت تصمم فيه دراسات أخرى، لمقارنة النهج القائم على المحتوى لاسترجاع المعلومات متعددة الوسائط، مقابل النهج القائم على الوصف (راجع على سبيل المثال **Petrelli & Auld, 2008**) وما لا يزال غائبا في هذا المجال الخاص ببحوث التقييم، هو الإطار، أو مجموعة المعايير التي يمكن أن تطبق للحكم على جودة نظم استرجاع المعلومات متعددة الوسائط، فأداء الاسترجاع فقط، على سبيل المثال، هو الذي يقاس بمعايير تقليدية كالتحقيق والاستدعاء، حتى في مسار تِرك الخاص بالفيديو **(Smeaton & Over, 2003) TREC Video track**، وما تلاه تركفيد **(Smeaton, 2007) TRECVID**. ولا يزال هناك بعض الجوانب الأخرى لنظم استرجاع معلومات الوسائط المتعددة، التي لم تخضع للاختبار في جهود التقييم.

وقد وضع الباحثون في استرجاع المعلومات الموسيقية (**MIR**)، وهو قسم واضح المعالم في استرجاع المعلومات متعددة الوسائط، منصتهم الخاصة بالتقييم، وذلك على غرار الإطار الأساس لمؤتمرات استرجاع النصوص

(Downie, 2003) TREC paradigm. وبناء على الوثائق التي جمعها داوني Downie (2004)، فإن تقييم استرجاع المعلومات الموسيقية MIR لا يزال في المرحلة الأولية لم يغادرها؛ لأن معظم المؤلفين لم يقدموا سوى مجرد بعض المناقشات النظرية للموضوع، دون إجراء أي تقييم فعلي. وبعبارة أخرى، لا تزال هناك حاجة لعمل الكثير لسد فراغ تقييم استرجاع المعلومات الموسيقية على وجه الخصوص، وتقييم استرجاع المعلومات متعددة الوسائط بوجه عام.

القابلية للإفادة معيارا للتقييم:

القابلية للإفادة usability مصطلح تم سكه منذ عهد قريب نسبيا، ليحل محل عبارة التعاطف مع المستفيد user-friendliness، التي جرت العادة على أن تكون الكلمة المفتاحية، التي عادة ما نجدها في سياق تفاعل الإنسان مع الحاسب (Rowley & Hartley, 2008)، على الرغم من إمكان تتبع جذوره في قانون مورز Mooers law لنظم استرجاع المعلومات، الذي اقترح عام ١٩٦٠ (راجع ١/ ١/ ٢/ ٣ في هذا الكتاب). ووفقا للمنظمة الدولية للتوحيد القياسي International Organization for Standardization (ISO, 1998) فإن القابلية للإفادة تعرف بوصفها"المدى الذي يمكن به الإفادة من منتج ما، من جانب مستفيدين محددين، لتحقيق أهداف محددة، بفاعلية، وكفاءة، وقناعة أو ارتياح، في سياق معين للإفادة".

وتوضح هذه المواصفة المعيارية الخاصة بالأيزو ISO، كيف يمكن التحقق من المعلومات اللازمة لأن توضع في الحسبان، عند التحقق من القابلية للإفادة أو تقييمها، من حيث أداء المنتج ورضاء المستفيد (Bevan, 2001). ومن ناحية أخرى، يقدم جاكوب نيلسن Jakob Nielsen، الذي حقق إسهامات مهمة لنشر القابلية للإفادة كنهج بديل لتقييم مواقع العنكبوتية – يقدم تعريفا بسيطاً للقابلية للإفادة:"القابلية للإفادة ... إنما هي في الأساس قضية ما إذا كان النظام صالحا بما فيه الكفاية لتلبية جميع

احتياجات ومتطلبات المستفيدين، وغيرهم من شركائهم المحتملين، كوسطاء المستفيدين والمديرين" (1993, P. 24).

ومما لاشك فيه أن كلا من تعريف الأيزو **ISO**، وتعريف نيلسن ينطويان على معانٍ أكثر اتساعا، بينما تدل القابلية للإفادة، في مجال تنظيم المعلومات واسترجاعها، في الأساس، على تقييم أداء نظم استرجاع المعلومات ورضاء المستفيد. وعلى هذا النحو، فإن القابلية للإفادة ليست مقياسا واحدًا لتقييم نظم استرجاع المعلومات، وإنما تشمل مجموعة من المعايير الخاصة بأهداف التقييم. ويقترح نيلسن **Nielsen (1993, 2003)** الخصائص الخمس التالية للقابلية للإفادة الجيدة:

- القابلية للتعلم: ما مدى السهولة بالنسبة للمستفيدين، في إنجاز المهام الأساس، التي يواجهونها في النظام لأول مرة؟
- الكفاءة: بمجرد أن يلم المستفيدون بالنظام، ما مدى السرعة التي يمكن بها إنجاز المهام؟
- القابلية للتذكر: عندما يرجع المستفيدون إلى النظام، بعد فترة من التوقف عن الإفادة منه، ما مدى السهولة التي يمكنهم بها استعادة الكفاءة؟
- الأخطاء: كم عدد الأخطاء التي يرتكبها المستفيدون، وما مدى جسامة هذه الأخطاء، وما مدى السهولة التي يمكن بها للمستفيدين التوقف عن ارتكاب الأخطاء؟
- الرضاء: ما مدى المتعة في الإفادة من النظام؟

وتردد هذه العناصر الخمسة بوضوح، صدى النقاط الخمس التي نوقشت في ١٠/٣/٣ في هذا الكتاب. ومن ناحية أخرى، يميل شنايدرمان وبليزانت **Schneiderman and Plaisant (2005)** للنظر إلى مصطلح القابلية للإفادة

usability بوصفه مرادفا للتصميم الجيد لواجهة التعامل good interface design، وذلك في الطبعة الرابعة من الكتاب الكلاسيكي "تصميم واجهة المستفيد؛ استراتيجيات للتفاعل الفعال بين الإنسان والحاسب "Designing the user interface: Strategies For Effective Human-Computer Interaction.

وقد استخدمت القابلية للإفادة، بصرف النظر عن طريقة تفسير دلالاتها، على نطاق واسع، كطريقة لتقييم جودة مختلف نظم استرجاع المعلومات؛ فقد تم على سبيل المثال، إجراء اختبار للقابلية للإفادة، على الفهارس المتاحة على الخط المباشر حديثة التطبيق (راجع على سبيل المثال Antelman, Lynema & Pace, 2006; Guha & Saraf, 2007)، والأنواع الأخرى من نظم استرجاع المعلومات. وفضلا عن ذلك، تم اختبار إحدى واجهات البحث (فلامنكو Flamenco) من جانب هيرست وآخرين (2002) .Hearst, et al، كما اختبر هاجا وكانيدا Haga and Kaneda (2005) نظاما لاسترجاع الفيديو، من وجهة نظر القابلية للإفادة. وقد أنشئت عدة مختبرات للقابلية للإفادة، منذ نهاية العقد الأخير من القرن العشرين، بتجهيزات مصممة خصوصًا، وبرمجيات وضعت في فضاء فيزيائي، صممت خصوصًا لذلك الهدف، حتى وإن كان ما تختبره هذه المختبرات بوجه عام، يتجاوز حدود القابلية للإفادة الخاصة بنظم استرجاع المعلومات.

مشروعات التقييم الكبرى لتنظيم المعلومات واسترجاعها:

مضى حوالي نصف قرن، منذ إجراء أول دراسة تقييم في استرجاع المعلومات عام ١٩٥٣ (Lancaster, 1979). وقد تم تنفيذ مشروعات عدة، لتقييم تنظيم المعلومات واسترجاعها، منذ ذلك الحين. وتجارب كرانفيلد (راجع على سبيل المثال Cleverdon, 1962)، ومشروع المدلرز MEDLARS (نظام تحليل واسترجاع

الإنتاج الفكري الطبي **MEDical Literature Analysis and Retrieval System)** (راجع على سبيل المثال **Lancaster, 1968**)، وتجربة سمارت **SMART**، نظام معالجة واسترجاع النصوص **(System for the Manipulation and Retrieval of Texts)** (راجع على سبيل المثال **Salton, 1981**)، ودراسة ستيرز **STAIRS (Storage and Retrieval System)** (راجع على سبيل المثال **Blair & Maron, 1985**)، وسلسلة مؤتمرات استرجاع النصوص **TREC** (راجع على سبيل المثال **Harman, 1993; Harman & Voorhess, 2006**)، فضلا عن بعض المشروعات الكبرى في هذه الفئة. ومن بين جميع دراسات التقييم الكبرى، سوف نتناول تجارب كرانفيلد، وسلسلة مؤتمرات استرجاع النصوص **TREC** بالتفصيل؛ نظرا لأنهما الأقوى تأثيرا في تاريخ تقييم تنظيم المعلومات واسترجاعها. كما نناقش أيضا المشكلات والقضايا المحيطة بمشروعات تقييم تنظيم المعلومات واسترجاعها.

تجارب كرانفيلد:

أجريت تجارب كرانفيلد على مرحلتين، هما كرانفيلد الأولى، وكرانفيلد الثانية، بين عامي ١٩٥٧ و١٩٦٧، وهما أول مشروعات التقييم المستفيضة، تم تنفيذهما بقيادة سيريل دبليو. كلفردون **C.W. Cleverdon**، الذي كان حينئذ اختصاصيا للمكتبات بكلية علوم الطيران بكرانفيلد في انجلترا **College of Aeronautics in Canfield, England**. وعلى الرغم من أن تجارب كرانفيلد كانت تتعلق بتقييم الطرق اليدوية لتنظيم المعلومات واسترجاعها، فإن المنهج الذي ابتكر في المشروع، قد استخدم فيما بعد لتقييم الكثير من الطرق الرقمية لتنظيم المعلومات واسترجاعها. وغالبا ما تسمى الدراسات المشابهة لتجارب كرانفيلد بالمشروعات المماثلة لكرانفيلد.

كرانفيلد الأولى:

صممت تجارب كرانفيلد الأولى لاختبار الكفاءة النسبية لأربعة نظم للتكشيف، في أداء الاسترجاع، في بيئة تنطوي على الكثير من المتغيرات الأخرى، كالمكشف، والوثيقة، واستفسار البحث. وقد استخدم الاستدعاء والتحقيق لقياس أداء الاسترجاع في هذا المشروع، كأول تطبيق من هذا النوع، في تجارب واسعة النطاق لتقييم تنظيم المعلومات واسترجاعها. يضاف إلى ذلك أنه كان من الممكن حساب الاستدعاء في كرانفيلد؛ نظرا لأن مجموعات الاختبار التي تكونت خصوصًا للمشروع كانت محصورة ومعروفة. وكانت نظم التكشيف الأربعة التي وقع عليها الاختيار لكرانفيلد الأولى هي:

- التصنيف العشري العالمي (**UDC**).
- فهرس موضوعي هجائي.
- خطة تصنيف وجهي.
- نظام المصطلح الواحد للتكشيف الترابطي **coordinate**.

تصميم التجربة:

وقع الاختيار على ثلاثة مكشفين للاختبار، وكانت لديهم فئات مختلفة من الخبرات والتآلف مع النظم التي يتم اختبارها؛ فقد كان لدى أحدهم معرفة موضوعية، ولدى الآخر خبرة بالتكشيف، بينما أتى الثالث(*) من إحدى المكتبات العامة مباشرة، ولم يكن لديه معرفة موضوعية أو خبرة في التكشيف. وقد تم الحصول

(*) قد لا يعلم البعض أن هذا الثالث، هو فردرك ولفرو لانكستر، الذي أصبح علماً يشار إليه بالبنان، في مجال استرجاع المعلومات. (المترجم)

على مئة وثيقة، موزعة بالتساوي، على المجال العام لعلوم الطيران، والمجال المتخصص وهو الديناميكا الهوائية عالية السرعة. وطلب من كل مكشف تكشيف كل من الوثائق المئة خمس مرات، وذلك بإنفاق دقيقتين، وأربع دقائق، وثماني دقائق، واثنتي عشرة دقيقة، وستة عشر دقيقة، في تكشيف الوثيقة. وكان متوسط عدد المداخل الكشفية التي تخصص لكل وثيقة، يتراوح ما بين اثنين وعشرة **(Cleverdon, 1962)**. وكانت استفسارات البحث الخاصة بالاختبار "مصنعة" إذ يقوم أعضاء من خارج المشروع، بفحص كل وثيقة، ثم صياغة الاستفسارات التي يمكن أن تكون هناك وثيقة معينة صالحة بالنسبة لها؛ فنشأ ما مجموعه ٤٠٠ استفسار للبحث، استخدمت في ثلاث دورات لعمليات البحث، لتشكل ١٢٠٠ (٤٠٠×٣) استفسارا.

وكانت عمليات البحث تتم بواسطة العاملين بالمشروع في المقام الأول. وقد أجريت دورة البحث الأولى علي المجموعة الأولى المكونة من ٤٠٠ استفسارا، كاختبار قبلي **pretest**. أما دورة البحث الثانية فكانت مصممة للتغلب على المشكلتين اللتين تمت مواجهتهما في الاختبار القبلي، وذلك باستخدام المجموعة الثانية المكونة من ٤٠٠ استفسارا، المشتقة من ٤٠٠ سؤالا. وكانت المشكلة الأولى هي إلى أي مدى كان من يقوم بإجراء عملية البحث مقتنعا بمواصلة البحث، بينما كان من المعروف أن الوثيقة الصالحة كانت في مكان ما بالملف؟ والثانية هي كيف يمكن لمن يقوم بإجراء عملية البحث، أن يقرر على وجه التحديد، ما كان يشكل استراتيجية بحث مختلفة، عند استعمال مصطلحات بعينها من لغات التكشيف الأربع؟ ثم تم بعد ذلك تحديد نقطة توقف بالنسبة لعملية البحث، في الدورة الثانية، وتطلب الأمر إدخال تعديلات على استراتيجيات البحث بالنسبة للاستفسارات نفسها، لتكون هذه الاستراتيجيات عامة، وبمصطلحات كل نظام من نظم التكشيف، وذلك على نحو محايد قدر الإمكان. وقد كشف تحليل إخفاق النتائج بالنسبة للجولة الثانية، عن حالات نجح

البحث فيها في أحد نظم التكشيف، إلا أن الاستفسار نفسه في نظام آخر أخفق نظرا لصياغة الاستراتيجية. ومن ثم، صممت دورة ثالثة للبحث باستعمال مجموعة أخرى من ٤٠٠ استفسارا، للتخلص من مظاهر التباين بين من يقومون بإجراء عمليات البحث واستراتيجيات البحث، وذلك باتباع استراتيجية معيارية ثابتة لكل النظم الأربعة.

نتائج الاختبار:

كانت أهم نتيجة انتهت إليها كرانفيلد الأولى، هي أن نظام المصطلح الواحد حقق أفضل استدعاء، من بين نظم التكشيف الأربعة المختلفة، التي شملها الاختبار (راجع الجدول رقم ١١/ ٢)، على الرغم من أن جميع النظم حققت مستوى مماثلا في كفاءة الأداء **(Cleverdon, 1962)**.

الجدول رقم ١١/ ٢ معدل استدعاء نظم التكشيف الأربعة

نظام التكشيف	الاستدعاء (٪)
المصطلح الواحد	٨٢.٠
الفهرس الموضوعي الهجائي	٨١.٥
التصنيف العشري العالمي	٧٥.٦
خطة التصنيف الوجهي	٧٣.٨

كذلك كشفت كرانفيلد الأولى عن أن تزايد الوقت المستنفد في التكشيف، أدى إلى تزايد الاستدعاء حتى مستوى معين، بعده يمكن للارتباط الإيجابي بين وقت التكشيف والاستدعاء، أن يصبح عكسيا، كما يتبين في الجدول رقم ١١/ ٣ **(Cleverdon, 1962)**.

الجدول رقم ١١/ ٣ زمن التكشيف في مقابل معدل الاستدعاء

زمن التكشيف (بالدقيقة)	الاستدعاء (٪)
٢	٧٢.٩
٤	٨٠.٢
٨	٧٦.٢
١٢	٨٢.٧
١٦	٨٤.٣

وبينما كان المكشفون الثلاثة، مختلفين من حيث خبرة التكشيف والمعرفة الموضوعية، فإنه لم يلاحظ اختلاف يذكر فيما بينهم، في استرجاعهم للوثائق التي كشفوها بأنفسهم، أو في أدائهم في التكشيف. وفضلا عن ذلك فإن معدلات النجاح في استرجاع الوثائق حول المجال العام لعلوم الطيران، لوحظ أنها أفضل بنسبة ٤ بالمئة و٥ بالمئة، من معدلات النجاح في استرجاع الوثائق حول المجال المتخصص للديناميكا الهوائية عالية السرعة. وتشمل النتائج الأخرى التي أسفرت عنها كرانفيلد الأولى ما يلي:

- هناك علاقة تناسب عكسي بين الاستدعاء والتحقيق.
- من الممكن تحقيق زيادة مقدارها واحد بالمئة في التحقيق على حساب فقد ٣ بالمئة في الاستدعاء.
- كانت النظم تعمل بمعدل استدعاء ما بين ٧٠ بالمئة و٩٠ بالمئة، ومعدل تحقيق ما بين ٨ بالمئة و٢٠ بالمئة.

ومن بين جميع النتائج التي لخصناها هنا، هناك ثلاث تتجاوز التوقعات، وربما تتناقض مع بعض الاعتقادات العامة السائدة في أوساط تنظيم المعلومات واسترجاعها. وأولى هذه النتائج الثلاث، فيما يتعلق بالاستدعاء، أن نظام المصطلح الواحد فاق في أدائه ثلاثة نظم للغات المقيدة شملها المشروع، على الرغم من الاعتقاد النظري بأن اللغة المقيدة تسفر عن أداء استرجاع أفضل من غيرها. وقد أدت هذه النتيجة بشكل غير مباشر، إلى الحث على البحث حول التكشيف الآلي، الذي يعتمد بقوة على المصطلحات الأحادية أو الكلمات المفتاحية. وثانيها، تبين أن أداء نظام التكشيف بالنسبة للاسترجاع، لا يتوقف على خبرة المكشف أو المعرفة الموضوعية، بينما كان الاعتقاد السائد على العكس تماما؛ أى إنه من الممكن لخبرة المكشف ومعرفته الموضوعية، أن يساعدا على تحسين الأداء. وثالثها، تبين من كرانفيلد الأولى أن الزيادة في زمن التكشيف، لا يمكن بالضرورة أن تؤدي إلى زيادة الاستدعاء. وعلى العكس، كان هناك تناقص في الاستدعاء عندما ارتفع زمن التكشيف من أربع دقائق إلى ثماني دقائق (راجع الجدول رقم ١١/ ٣). وبينما ارتفع معدل الاستدعاء ثانية، نتيجة لزيادة زمن التكشيف، فإن العائد في أداء الاسترجاع لا يبدو متناسبا مع الوقت المستنفد في التكشيف؛ فقد ارتفع معدل الاستدعاء، على سبيل المثال، بمقدار ٤ بالمئة فقط (أي من ٨٠.٢ بالمئة إلى ٨٤.٣ بالمئة)، بينما تضاعف الوقت المستنفد في التكشيف أربع مرات (أي من أربع دقائق إلى ست عشرة دقيقة).

كرانفيلد الثانية:

لقد صممت تجارب كرانفيلد الثانية لتقييم تأثير مختلف أدوات التكشيف على أداء الاسترجاع. وأدوات التكشيف هي التدابير التي يتخذها المكشف للارتفاع بمستوى الاستدعاء أو التحقيق. وتشمل أمثلة أدوات التكشيف المترادفات، وعلاقة العام بالخاص **generic relations**، والربط **coordination**، والروابط **links**،

والأدوار **roles**. وقد وردت المعلومات حول التجارب في تقرير كلفردون عن المشروع **(Cleverdon, 1962)**:

> لقد انطلقنا من الاعتقاد بأن جميع لغات التكشيف مزيج من فئات مختلفة من الأدوات. وتنقسم مثل هذه الأدوات إلى فئتين؛ تلك الأدوات التي يقصد بها تحسين معدل الاستدعاء، وتلك التي يقصد بها تحسين معدل التحقيق... وكان هدف التجربة دراسة ما يمكن أن يكون لكل من هذه الأدوات، منفردة أو في أي تجمع محتمل، من أثر على الاستدعاء والتحقيق. **(P. 17)**

تصميم التجربة:

تم تجميع نحو مئتي بحث، بكل بحث بوجه عام، ما بين خمسة إلى عشرة مراجع. ثم طلب بعد ذلك من كل مؤلف أن يبين، في شكل تساؤل بحث، المشكلة الرئيسة التي كانت سببا لدراسته، على افتراض أن البحوث قد كتبت نتيجة لدراسات أجريت لتقديم إجابات لتساؤل ما، أو عدة تساؤلات في ذهن المؤلف. كذلك طلب من كل مؤلف أن يصوغ في حدود ثلاثة تساؤلات إضافية، تعبر عن أي قضايا فرعية أثيرت أثناء إعداد الدراسة. وبعد ذلك تم الحصول على مستخلصات لمراجع هذه البحوث (أي الوثائق المستشهد بها، إلى جانب بعض البحوث غير المستشهد بها، التي تم التحقق منها من جانب طلبة الدراسات العليا بالكلية، بناء على العناوين، أو عن طريق المزاوجة الوراقية **bibliographical coupling** (إذ تتزاوج وثيقتان وراقيا عندما تشتمل قوائم مراجعها على وثيقة واحدة مشتركة أو أكثر). وقد طلب من المؤلفين الكشف عن صلاحية المراجع والبحوث غير المستشهد بها، بالنسبة لكل تساؤل، على مدرج من واحد إلى خمسة، حيث يدل الواحد على الأكثر صلاحية

على الإطلاق، وذلك على افتراض أن المراجع قد سجلت لأنها ينبغي أن تكون متصلة على نحو ما، بموضوع البحث (Cleverdon & Mills, 1963). وقد تحقق المؤلفون من ١٩٦١ وثيقة مستشهد بها، صالحة كليا أو جزئيا للأسئلة التي طرحوها، البالغ عددها ٢٧٩ سؤالا، وقد تم اختيار ١٤٠٠ وثيقة و٢٢١ سؤالا من هذه المجموعة، في النهاية لكرانفيلد الثانية.

وكانت المجموعة التجريبية المكونة من ١٤٠٠ وثيقة، في المقام الأول، في مجال الديناميكا الهوائية عالية السرعة وتصميم الطائرات. وقد "كشفت" كل وثيقة بثلاث طرق؛ تم في الأولى اختيار أهم المفاهيم وتسجيلها باللغة الطبيعية للوثيقة، وفي الثانية تم تسجيل المصطلحات التي وردت في كل من المفاهيم التي وقع عليها الاختيار، وفي الثالثة تم الجمع بين المفاهيم بطرق مختلفة، لتشكل الموضوعات الرئيسة للوثائق. كذلك أعطي لكل مصطلح وزن (أي واحد أو اثنان أو ثلاثة، حيث واحد هو الأكثر أهمية على الإطلاق) وذلك في أثناء عملية التكشيف، للدلالة على أهميته النسبية. وقد استخدمت أدوات تكشيف مختلفة أيضا أثناء عملية التكشيف، من أجل استكشاف مالها من تأثير على أداء الاسترجاع.

وقد تمت عمليات البحث باستخدام مختلف أدوات التكشيف المطبقة في ثلاثة أنواع رئيسة للغات التكشيف، وهي لغات تكشيف المصطلح الأحادي، والمفهوم المفرد، والمصطلح المقيد أو المقنن. وقد وردت المواصفات الخاصة بلغات التكشيف هذه، تفصيلا في الشكل رقم (٣) والشكل رقم (٥) في مقالة كلفردون Cleverdon (1967). كذلك تم اختبار لغات التكشيف المكونة من الكلمات المفتاحية الواردة بالعناوين والمستخلصات، أيضا في المشروع. وكانت الاستفسارات تحلل إلى مصطلحات مفردة، وكانت عملية البحث تتم أولا للبحث عن مضاهاة بالنسبة لكل المصطلحات، وكانت المضاهاة تتم بعد ذلك على مستويات مختلفة،

وذلك باستبعاد أحد مصطلحات الاستفسار في كل مرة، وبالنسبة لكل ترابط محتمل بالمصطلحات المتبقية؛ فبالنسبة لاستفسار به خمسة مصطلحات مفردة على سبيل المثال، كانت عملية البحث الأولى تتم بكل المصطلحات الخمسة مجتمعة. وكان من الممكن إجراء عملية البحث الثانية بأي من المصطلحات الأربعة من الخمسة، باستبعاد مصطلح واحد من الاستفسار. وبعد ذلك تتم عملية البحث بأي من المصطلحات الثلاثة، ثم بأي من المصطلحين، إلى أن يصبح هناك مصطلح واحد للمضاهاة. وكان من الممكن استعمال المترادفات، والمصطلحات الأضيق والمصطلحات الأوسع، والمقومات الأخرى، لتحل محل الكلمات الأصلية. وكانت عمليات البحث تتم على نحو يكفل المحافظة على جميع العوامل ثابتة، فيما عدا عامل واحد متغير. وفي المثال السابق، كان عدد المصطلحات الواردة في الاستفسار هو العامل الوحيد المسموح بتغيره، بينما كانت تتم المحافظة على جميع العوامل الأخرى ثابتة. وكانت كل من التسديدات **hits**، والشوشرة **false drops**، أي الوثائق غير الصالحة المسترجعة، يتم تسجيلها بالنسبة لكل عملية بحث، بحيث يمكن حساب معدلات الاستدعاء والتحقيق.

وكان التصميم برمته، لافتا للنظر نظرا لحجمه، وتنوعه، وتعقده. ولم يكن الاختبار على أي حالٍ سهلاً، وخصوصا إذا علمنا أنه قد تم إجراؤه يدويا.

نتائج التجربة:

لقد تم حساب متوسط معدلات الاستدعاء والتحقيق الخاصة بجميع عمليات البحث، في كل لغة من اللغات، واستخدم هذا المتوسط لقياس تأثير أدوات التكشيف على أداء الاسترجاع في كرانفيلد الثانية. وقد سجلت التجربة النتائج التالية:

- كانت لغات التكشيف المعتمدة على المصطلحات المفردة، متفوقة من حيث الأداء، على أي نوع آخر.
- عندما استخدمت المصطلحات المفردة للتكشيف، أدى إدخال الفئات المقابلة، وخصوصا أشباه المترادفات، إلى انخفاض مستوى الأداء.
- عندما استخدمت المفاهيم للتكشيف، تراجع مستوى الأداء، بإدخال الفئات الرئيسة والفرعية والمناظرة، إلى جانب المفاهيم الأصلية.
- عندما استخدمت المصطلحات المقيدة للتكشيف، أدى إدخال المصطلحات الأضيق والمصطلحات الأوسع، إلى تراجع مستوى الأداء.
- كان أداء لغات التكشيف، المشكلة من العناوين، أفضل من أداء تلك المشكلة من المستخلصات.

وكان ما يدعو للعجب فعلا، ولا يمكن تفسيره ببساطة، أنه كما تبين في كرانفيلد الثانية، أن أفضل لغات التكشيف أداء، كانت تتكون من كلمات مفردة غير مقيدة، منتقاة من الوثائق. ولم يكن بإمكان المشاركين في البحث، التوصل إلى تفسيرات حول هذه المفاجأة **(Cleverdon, Mills & Keen, 1966)**. وتتطابق هذه النتيجة أيضاً مع إحدى نتائج كرانفيلد الأولى، وهي أن نظام المصطلح الواحد **uniterm** قد تفوق، في أدائه على نظم اللغات المقيدة الثلاثة الأخرى، التي شملها المشروع، من حيث الاستدعاء. وبناء على كرانفيلد الثانية، فإنه يبدو أن المصطلحات الأصلية تؤدي دائما، على نحو أفضل وحدها، دون أي تنويعات إضافية (كالمصطلحات الأوسع والمصطلحات الأضيق على سبيل المثال). كذلك قدمت لغات التكشيف المعتمدة على العناوين، أداء أفضل من أداء اللغات المعتمدة على المستخلصات؛ نظرا لأن العناوين بوجه عام، أكثر دلالة ودقة من المستخلصات. كذلك تأكدت أيضا علاقة التناسب العكسي بين الاستدعاء والتحقيق، في كرانفيلد الثانية، عندما أصبحت أدوات التكشيف جزءا من البيئة التجريبية.

مشكلات تجارب كرانفيلد:

نشر عدد كبير من المراجعات والانتقادات (راجع على سبيل المثال **Harter, 1971; Swanson, D.R., 1965, 1971; Vickery, 1966, 1967**) منذ أجريت تجارب كرانفيلد، تشير إلى مختلف المشكلات التي اكتنفت هذه التجارب. وقد صنفت هارتر وهيرت **Harter and Hert (1997)** هذه المشكلات في أربع فئات: الصلاحية وإمكان الاعتماد أو التعويل **validity and reliability**، والقابلية للتعميم **generalizability**، والفائدة أو الجدوى **usefulness**، والقضايا النظرية **conceptual issues**. ونظراً لأن القضايا النظرية أو المفاهيمية (مثل كيف يمكن تعريف وقياس مفاهيم كالصلاحية والرضاء) موجودة في كل مشروعات تقييم تنظيم المعلومات واسترجاعها؛ فإن المناقشة التالية سوف تدور فقط حول ثلاث فئات من المشكلات، المتصلة على وجه التحديد بتجارب كرانفيلد، وهي الصلاحية وإمكان الاعتماد عليها، والقابلية للتعميم، والقابلية للتطبيق.

ولقد كانت صلاحية التجارب وقابليتها للاعتماد عليها، موضوعا للتساؤل نظرا لعدة حجج؛ فمن الممكن، على سبيل المثال، للوثائق التي قدمت فعلا للمؤلفين الذين طرحوا الأسئلة، لتقييم الصلاحية فيما يتعلق بالأسئلة المقابلة، أن تكون في صالح أساليب مضاهاة المصطلحات البسيطة، على حساب بعض نظم المضاهاة المركبة. كما لم تدخل أعداد كبيرة من الوثائق، التي كان من المحتمل أن تكون صالحة ولكنها لم تسترجع، في التقييم على الإطلاق. إلا أن النظام لم يحكم عليه سلبا؛ نظراً لعدم استرجاع هذه الوثائق. وكانت هناك أيضا افتراضات ليس لها ما يبررها، أو لم تخضع للاختبار في التجارب؛ فعلى المرء على سبيل المثال، أن يفترض أن تكون أحكام الصلاحية التي صدرت على عدد من الوثائق، بالنسبة لسؤال معين، مناقضة لبعضها البعض. يضاف إلى ذلك أن الارتباط بين الوثائق الأصلية والأسئلة، في كرانفيلد الثانية، كان أقوى بكثير مما هو عليه في كرانفيلد الأولى. ويؤثر كل ذلك سلباً في

صلاحية تجارب كرانفيلد، وإمكان التعويل عليها؛ أي ما إذا كانت التجارب قد قاست فعلاً، ما كان قد قصد بها قياسه، وما إذا كانت التجارب قابلة للتكرار، إذا ما اتبعت الإجراءات نفسها.

وتبدو نتائج كرانفيلد غير قابلة للتعميم؛ لأنه لم تستخدم العينة العشوائية في التجارب. وفضلا عن ذلك، كانت المجموعات التجريبية صغيرة (أي مئة وثيقة في الأولى و١٤٠٠ وثيقة في الثانية) بالمقارنة بالعدد الفعلي للوثائق في نظام ما لاختزان المعلومات واسترجاعها. وتبدو بعض النتائج، كما سبق أن بينا، مناقضة للاعتقاد العام، والخبرة المكتسبة في أوساط تنظيم المعلومات واسترجاعها. وتقف كل هذه الحقائق وراء المخاوف بشأن قابلية هذه التجارب للتعميم.

كذلك تبدو قابلية التجارب للتطبيق في النظم العاملة، موضوع تساؤل، نظرا للاصطناعية البادية في عدة أوجه لتجارب كرانفيلد؛ أولها أنه لم تكن هناك مشاركة حقيقية للمستفيدين في جميع مراحل التجارب. ثانيا، كانت أسئلة البحث مصنعة، ولم تكن موجهة من جانب مستفيدين حقيقيين. ثالثا، لم يكن نظام الاختبار نظاما عاملا، وإنما صمم لمجرد إجراء التجارب. ومما لاشك فيه، أنه من المستحيل التحكم في متغيرات الاختبار (كعدد المصطلحات التي ترد في الاستفسار، والوقت المستنفد في استرجاع الوثائق، على سبيل المثال) التي ينطوي عليها الموقف في نظام عامل فعلا؛ ومن ثم، فإن ما تم التوصل إليه في بيئة تجارب كرانفيلد الاصطناعية، قد لا يكون قابلا للتطبيق على النظم العاملة لاختزان المعلومات واسترجاعها.

أهمية تجارب كرانفيلد:

على الرغم من كل ما وجه من انتقادات، فقد أحدثت تجارب كرانفيلد تأثيرا ضخما في كثير من مشروعات التقييم الأخرى، على نحو ما يلي من أوجه؛ أولا تحققت تجارب كرانفيلد من العوامل الرئيسة التي تؤثر في أداء نظم الاسترجاع. ولغات

التكشيف وأدوات التكشيف مجرد أمثلة لهذه العوامل. ثانيا طورت تجارب كرانفيلد، لأول مرة منهجا لا يزال متبعاً في مؤتمرات استرجاع النصوص **TREC**، على الرغم من عيوبه الكثيرة. ثالثا أثرت تجارب كرانفيلد بوضوح، في نظرتنا لنظم استرجاع المعلومات، وكيف ينبغي دراستها، لما فيه الصالح العام على الإطلاق تقريبا **(Spärck Jones, 1981)**. رابعا، مهدت تجارب كرانفيلد الطريق نحو أساليب التكشيف الآلي الفعالة القابلة للتحقيق؛ نظرا لأنها أثبتت أن نظم المصطلح الواحد واللغة الطبيعية، تميل لأن يكون أداؤها مماثلا، إن لم يكن أفضل من أداء نظم اللغات المقيدة.

ولا يمكن المبالغة في أهمية تجارب كرانفيلد، حتى وإن بدت مشكلاتها لا يستهان بها. وفيما يتعلق بتقييم نظم اختزان المعلومات واسترجاعها، فإن جميع المشروعات الكبرى، بما في ذلك سلسلة مؤتمرات استرجاع النصوص **TREC**، تبدو متأثرة بطريقة أو أخرى بتجارب كرانفيلد. ولا يمكن لمشروع تقييم واسع المدى بعينه، أن يصبح استثناءً، في مجال تنظيم المعلومات واسترجاعها. وذلك بعينه هو السبب وراء سك مصطلح "المشروعات على غرار كرانفيلد **Cranfield-like projects**(*)".

(*) مما لا شك فيه أن دراسات كرانفيلد قد أرست أسس المنهج التجريبي في مجال علم المعلومات.

راجع فى ذلك: حشمت قاسم. دراسات كرانفيلد وتطور مناهج البحث في علم المعلومات. مجلة المكتبات والمعلومات العربية، مج١، ع٤؛ أكتوبر ١٩٨١. ص ص ٤٩-٩٥. وأعيد نشر هذه المقالة في:

حشمت قاسم: دراسات فى علم المعلومات، ط٢. القاهرة، دار غريب، ١٩٩٥. ص ص ٢١١-٢٤٩. (المترجم)

سلسلة مؤتمرات استرجاع النصوص:

تشكل سلسلة مؤتمرات استرجاع النصوص **TREC series**، استمرارا لمؤتمرات كانت تعقد في الولايات المتحدة الأمريكية سنويا، منذ عام ١٩٩٢ (أي ترِك-١ **TREC-1**)، بقصد تحقيق الأهداف التالية:

- تشجيع البحث في استرجاع النصوص، اعتمادا على مجموعات ضخمة من البيانات النصية.
- دعم التواصل بين القطاع التطبيقي، والأوساط الأكاديمية، والحكومية، بتوفير منتدى منفتح لتبادل أفكار البحوث.
- زيادة سرعة نقل التقنيات من مختبرات البحث إلى المنتجات التجارية، باستعراض التحسينات الجوهرية في تقنيات الاسترجاع المتعلقة بمشكلات الواقع.
- التوسع في إتاحة أساليب التقييم الملائمة للاستخدام من جانب القطاع التطبيقي، والأوساط الأكاديمية، بما في ذلك تطوير أساليب التقييم الجديدة، القابلة أكثر من غيرها للتطبيق على النظم الحالية.

وكانت هذه المؤتمرات تحظى بالرعاية المشتركة بين المعهد الوطني للمواصفات المعيارية والتقنيات (نست **National Institute of Standards and Technology** **NIST**)، وجهاز مشروعات بحوث الدفاع المتطورة **Defense Advanced Research Projects Agency (DARPA** (داربا**)**، والأجهزة والمؤسسات الأخرى، ذات العلاقة (مثل أنشطة البحوث المتطورة والبحوث والتطوير **Advanced Research and Development Activity (ARDA** (أردا**)**، وديوان وزارة الدفاع **Department of Defense**) في الولايات المتحدة الأمريكية.

وبنهاية عام ٢٠٠٨، كان قد عقد ١٧ مؤتمرا لترِك TREC. وعلى الرغم من تغير التفصيلات حول سلسلة ترِك عبر السنين، كالمشاركين والمسارات (أي المشكلات الفرعية في استرجاع المعلومات)، ظلت أهداف أكبر جهد في التقييم في تاريخ تنظيم المعلومات واسترجاعها، كما هي.

تصميم سلسلة مؤتمرات استرجاع النصوص:

سلسلة مؤتمرات استرجاع النصوص **TREC series**، على عكس تجارب كرانفيلد، مشروع تقييم مستمر، يضم فرقا مختلفة من المشاركين. وتقوم بعض هذه الفرق بتطوير النظم الخاصة بها لاسترجاع المعلومات، ودائما ما تحصل الفرق التي تقرر المشاركة في مؤتمر بعينه، على المجموعة التجريبية، والمعلومات المتصلة بها، التي يقدمها رعاة المؤتمر ومنسقو المسارات. ويتم الاسترجاع الفعلي في المقار الخاصة بالفرق المشاركة، وتقدم نتائج التقييم المتربعة على القمة للرعاة من أجل التقييم. كذلك تجتمع الفرق المشاركة في كل مؤتمر لاسترجاع النصوص، إبان الاختبار القبلي **pretest**، لتبادل وجهات النظر والخبرات، ووضع الخطط للمؤتمر التالي لاسترجاع النصوص.

الفرق المشاركة:

تمثل الفرق المشاركة في مؤتمرات استرجاع النصوص (مثل بلكور، جامعة كارنيجي ملون **(Belcore, Carnegie Mellon University** القطاع التطبيقي، والأوساط الأكاديمية، والمؤسسات الحكومية في الولايات المتحدة الأمريكية، ومن الدول الأخرى أيضا (كالصين وفرنسا على سبيل المثال). وكانت المشاركة في مؤتمرات استرجاع النصوص تتم في البداية بالدعوة فقط، إلا أنها الآن متاحة لكل من لديه القدرة على المشاركة في المشروع. ويشتمل الجدول رقم ١١/ ٤ على أعداد الفرق المشاركة في المؤتمرات التسعة الأولى لاسترجاع النصوص.

وبإمكان الفرق المشاركة الاختيار من بين ثلاثة مستويات للمشاركة؛ المشاركة الكاملة، والمشاركة الكاملة باستخدام ربع مجموعة الوثائق التجريبية الكاملة، والمشاركة في التقييم فقط، وذلك لكفالة حماية النظم التجارية لأسرار الخوارزميات التي تمتلكها (**Harman, 1995**).

الجدول رقم ١١/ ٤ عدد الفرق المشاركة في مؤتمرات استرجاع النصوص

مؤتمر استرجاع النصوص	العام	عدد الفرق
المؤتمر الأول	١٩٩٢	٢٥
المؤتمر الثاني	١٩٩٣	٣١
المؤتمر الثالث	١٩٩٤	٣٣
المؤتمر الرابع	١٩٩٥	٣٦
المؤتمر الخامس	١٩٩٦	٣٨
المؤتمر السادس	١٩٩٧	٥١
المؤتمر السابع	١٩٩٨	٥٦
المؤتمر الثامن	١٩٩٩	٦٦
المؤتمر التاسع	٢٠٠٠	٦٩

الوثائق التجريبية:

لقد تم تجميع الوثائق التجريبية، في المقام الأول، من الصحف (مثل **Wall Street Journal**)، والشبكات الإخبارية **newswires** (مثل شبكة أسيوشيتدبرس الإخبارية **Associated Press Newswire**)، وأنواع أخرى من المصادر (مثل السجل الاتحادي

Federal Register)[*]. وكانت هذه الوثائق تصاغ بشكل موحد بلغة الترميز العامة المعيارية Standard Generalized Markup Language (SGML)، وتوزع على أسطوانات مكتنزة، سعة كل منها واحد جيجا بايت من البيانات. وقد بدأ المؤتمر الأول بأسطوانتين للبيانات التجريبية، ثم ارتفع عدد الأسطوانات المستخدمة في التجارب، في النهاية، إلى خمس. وتشمل الأسطوانتان ٤ و٥ على بيانات خاصة بالمهمة الخاصة **ad hoc task** التي نتناولها فيما بعد. وتتراوح أطوال الوثائق ما بين ٣٠٠-٤٠٠ مصطلح، وعدة مئات من الصفحات. وتتكون المجموعة التجريبية من أكثر من مليون وثيقة.

وقد أدى توسع مؤتمرات استرجاع النصوص في مسارات متعددة، إلى تصميم وبناء كثير من مجموعات النصوص المتخصصة. وبالإضافة إلى تلك الوثائق الواردة في الجدول رقم ٤/ ١ في المراجعة العلمية التي أعدها هارمان وفورهيس **Harman and Voorhees (2006)**، جمعت وثائق تجريبية أخرى، لمسارات كالجينوم **Genome**، والجوانب القانونية **Legal**. وبعبارة أخرى، فإن المجموعات التجريبية المستخدمة في المسارات الحالية، لم تعد كتلك التي تكونت للمسارات المبكرة، من حيث المحتوى والشكل والمجال.

الموضوعات والاستفسارات:

لقد كانت الموضوعات[**] الخاصة بالاسترجاع، يكتبها أناس ممن كانوا يفيدون فعلا من نظام الاسترجاع، وكانت مهيأة لتحاكي احتياجات المستفيدين الحقيقية. وكانت الموضوعات تبنى بإجراء عمليات استرجاع تجريبية، في مقابل عينة من مجموعة الوثائق، وبعد ذلك يتم اختيار تلك الموضوعات التي تحرز ما بين ٢٥ إلى ١٠٠

(*) الصحيفة اليومية الرسمية للولايات المتحدة الأمريكية. (المترجم)

(**) تعني هنا الاستفسارات. (المترجم)

تسديدة **hit**، على وجه التقريب، في تلك العينة. وكان كل موضوع يصاغ بالطريقة المعيارية نفسها (أي بلغة الترميز العامة المعيارية **SGML**، مع رقم تسلسلي، وعنوان قصير، ووصف مكون من جملة واحدة، وما شابه ذلك)، لكفالة ترتيب آلي أكثر سهولة للاستفسارات. وكانت الموضوعات الخمسون نفسها، تعد لمهمة التوجيه في كل مؤتمر من مؤتمرات استرجاع النصوص **TREC**، مع تكوين ٥٠ موضوعا جديدا، لكل مؤتمر من أجل المهمة الخاصة. وسوف نوضح بعد قليل ما يقصد بالمهمة الخاصة، ومهمة توجيه الاسترجاع الخاصة بمؤتمرات استرجاع النصوص. كذلك كان يتم أيضا إعداد موضوعات للتدريب، لكفالة تعرف المشاركين على مجموعة البيانات التجريبية. وكانت جودة الموضوعات بالنسبة للمؤتمرين الأول والثاني عالية جدا، إلى حد أنه كان من الممكن بلوغ مستوى جيد جدا للأداء، دون استعمال استراتيجيات التكشيف والبحث المرهقة بلا مبرر؛ ومن ثم، فقد تم إعداد الموضوعات، بدءا من المؤتمر الثالث فصاعدا، واضعين في الحسبان قدرا كبيرا من الواقعية، في الوقت الذي كانت تراعي فيه الاهتمامات التشخيصية في التقييم المختبري أو التجريبي **(Spärck Jones, 2000)**.

وكانت الموضوعات تقدم كبيانات بالاحتياجات، لا كطلبات تقليدية، بناء على ثلاثة اعتبارات **(Voorhees & Harman, 2005, pp. 28-29)**. وأول هذه الاعتبارات أن يكون بإمكان الموضوعات التي تعد في شكل بيانات احتياجات، كفالة مدى واسع من طرق صياغة الاستفسارات. أما الاعتبار الثاني فهو أن يكون بإمكان الموضوعات بهذا الشكل، توفير المزيد من المعلومات لكل من المسئول عن إجراء عملية البحث ومن يقوم بالتقييم، يفيدان منها في إصدار أحكام الصلاحية. والاعتبار الثالث، أن يكون بإمكان الموضوعات من هذا النوع، تيسير إعادة استعمال المجموعة التجريبية نفسها، في المستقبل، لتجارب مختلفة؛ نظرًا لأن المعلومات التفصيلية التي

تتضمنها الموضوعات جزء من المجموعة. ثم يتم بعد ذلك صياغة استفسارات الاسترجاع، بناء على الموضوعات، آليا أو يدويا. وكان من الممكن للطريقة اليدوية أن تستخدم بقدر من المساعدة الآلية. وكانت بعض الاستفسارات الخاصة، تصاغ باستعمال التلقيم المرتد للصلاحية **relevance feedback**.

مهام الاسترجاع:

لقد تم في مؤتمرات استرجاع النصوص **TERC** اختبار نوعين من مهام الاسترجاع، وهما المهمة الخاصة **ad hoc**، والتوجيه **routing**، وكانت المهام الخاصة تقابل الاسترجاع الراجع **retrospective**، أي العثور على كل المعلومات حول موضوع معين، بينما كانت مهمة الاسترجاع بالتوجيه تضاهي البث الانتقائي للمعلومات، بالعثور على المعلومات التي أنتجت حديثا حول استفسار ما.

وبدءا من المؤتمر الرابع، أدخلت مجموعة من مهام الاسترجاع الإضافية، تركز كل واحدة منها على مشكلة استرجاع فرعية معينة. وقد أوضحت سبارك جونز **(2000) Spärck Jones** ما يتعلق بالمسارات بصياغة التناظر الملائم التالي:

> من الممكن النظر إلى مؤتمرات استرجاع النصوص برمتها، على نحو غير رسمي، بوصفها دولاب **Wheel** بمحور مركزي (المهمة الخاصة **ad hoc**) ومكابح دوارة (المسارات)، ترتبط كل منها بالمحور، وتسلم الإطار بعضها لبعض. ومن الممكن إدخال المكابح الجديدة، وتقوية المكابح القائمة، أو نزع المكابح البالية. وبذلك يمكننا البحث عن دولاب استرجاع للمعلومات في ترِك، دائم التحسن، وليس بمجرد كوابح أكثر وأفضل فعلا، وإنما بإعادة تصميم المحور والإطار وتقويتها أيضا. **(pp. 43-44)**

ويشتمل الجدول رقم ١١/ ٥ على المسارات التي تمت تغطيتها، بدءا من المؤتمر الرابع حتى المؤتمر العاشر، الذي يعرف أيضا بمؤتمر ٢٠٠١. والواقع أن مؤتمرات استرجاع النصوص، منذئذ كانت تميز بالعام على نحو مطرد (ترِك ٢٠٠٢، وترِك ٢٠٠٩ على سبيل المثال) بدلا من خطة الترقيم السابقة. ويشتمل الجدول رقم ١١/ ٦ على المسارات التي تم إجراء التجارب فيها، من ترِك ٢٠٠٢ حتى ترك ٢٠٠٩. واعترافا بوجود بنية أساس كافية لدعم الباحثين المهتمين بمهام الاسترجاع الراجع، توقفت المهمة الخاصة في ترِك -٩، بحيث أصبح من الممكن تخصيص المزيد من موارد ترِك لإقامة بنية أساس لتقييم المسارات **(Voorhees, 2000)**.

الجدول رقم ١١/ ٥ نظرة عامة على مسارات ترِك: ترِك-٤ حتى ترِك-١٠

المسارات	ترِك-٤	ترِك-٥	ترِك-٦	ترِك-٧	ترِك-٨	ترِك-٩	ترِك-١٠
التفاعلي	×	×	×	×	×	×	×
ادماج مراصد البيانات	×	×					
متعدد اللغات	الإسبانية	الإسبانية الصينية	الصينية				
التشويش	×	×					
اللاسلكي المحدد بالبرمجيات SDR(*)			×	×	×	×	

(*) نظام للاتصالات اللاسلكية. (المترجم)

التنقية	×	×	×	×	×	×	×
معالجة اللغة الطبيعية NLP		×	×				
جهاز تشغيل الوسائط VLC		×	×	×			
استرجاع المعلومات عبر اللغات CLIR			×	×	×	×	×
التحقيق المرتفع			×	×			
الاستفسار				×	×	×	
إجابة الأسئلة					×	×	×
العنكبوتية					×	×	×
الفيديو							×

ونظرا لاندماج مهمة التوجيه مع مسار التنقية، منذ تـرِك-٧، أصـبحت جميـع مهام الاسترجاع التي تم استكشافها في مؤتمرات استرجاع النصوص، بدءا من ترِك-٧ تتخذ شكل المسارات.

الجدول رقم ١١/ ٦ نظرة عامة على مسارات ترِك: ترِك ٢٠٠٢ حتى ترك ٢٠٠٩

المسارات	ترِك ٢٠٠٢	ترِك ٢٠٠٣	ترِك ٢٠٠٤	ترِك ٢٠٠٥	ترِك ٢٠٠٦	ترِك ٢٠٠٧	ترِك ٢٠٠٨	ترِك ٢٠٠٩
التفاعلي	×							
التنقية	×							
استرجاع المعلومات عبر اللغات CLIR	×	×						
إجابة الأسئلة	×	×	×	×	×	×		
العنكبوتية	×	×	×					×
الفيديو/ ترك فيد	×	ترك فيد	×	×	×	×	×	×
الجدة	×	×	×					
دراسات الجينوم		×	×	×	×	×		
الاسترجاع عالي الدقة من الوثائق		×	×	×				
الاسترجاع		×	×	×				

السليم								
التيرابايت Terabyte			×	×	×			
المشروع				×	×	×	×	
البريد المزعج				×	×	×		
التدوين					×	×	×	×
القانوني					×	×	×	×
المليون استفسار						×	×	×
تلقيم الصلاحية المرتد							×	×
استرجاع المعلومات الكيميائية							×	×
الكيانات								×

وبحلول عام ٢٠٠٩ كان قد نشأ أكثر من عشرين مسارا، الأمر الذي يدل على نقاط تركيز بحوث أوساط استرجاع المعلومات في ذلك الوقت. وبينما توقفت بعض المسارات، أنشئت مسارات جديدة لدراسة المشكلات الناشئة في استرجاع المعلومات. واعتمادا على النظرات الشاملة، والانطباعات التي حظيت بها مؤتمرات استرجاع

النصوص التي عقدت (راجع على سبيل المثال **Harman, 1995; Harman & Voorhees, 2006; Spärck Jones, 1995, 2000; Voorhess & Harman, 2005**)، وفيما يلي وصف موجز لكل مسار، في ترتيب هجائي(*):

- مسار المدونات **Blog track**، بدأ في مؤتمر ٢٠٠٦، لاستكشاف سلوكيات البحث عن المعلومات، في مجال المدونات. وكان هذا المسار لا يزال جاريا في مؤتمر ٢٠٠٩.

- مسار استرجاع المعلومات الكيميائية **Chemical IR track**، أحد مسارين جديدين، أضيفا في مؤتمر ٢٠٠٩ (إلى جانب مسار الكيانات **Entity**)، ويسعى إلى تطوير وتقييم التقنيات الخاصة بإجراء عمليات البحث واسعة المدى، في الوثائق الكيميائية، بما في ذلك البحوث الأكاديمية وبراءات الاختراع. والمستفيدون المستهدفون من هذا المسار هم اختصاصيو عمليات البحث عن المعلومات، وخصوصاً باحثو براءات الاختراع والكيميائيون.

- مسار التشويش **confusion track**، وكان يمثل الجهود التي تبذل مع البيانات المحرفة **corrupted**، التي يمكن أن تتأتى، على سبيل المثال، نتيجة للتحقق من الأحرف بصريا (**OCR**)، أو المدخلات الناطقة أو المسموعة. وتسمى بيانات ذلك الطابع أيضا بالبيانات المشوشة أو المحرفة. وقد تم تطبيق هذا المسار في المؤتمرين الرابع والخامس، وهو يتصل بالبحث المشوش **fuzzy** الذي نوقش في ٥/ ١/ ٢/ ١ في هذا الكتاب.

- مسار استرجاع المعلومات عبر اللغات **Cross-Language Information Retrieval (CLIR) track**، وهو مسار تغطي فيه الوثائق عدة لغات

(*) وفقاً لأسمائها باللغة الإنجليزية بالطبع. (المترجم)

(كالإنجليزية، أو الفرنسية، أو الألمانية، أو الإيطالية على سبيل المثال) وتغطى الموضوعات بكل لغة من هذه اللغات. ويركز هذا المسار على استرجاع الوثائق بإحدى اللغات المستهدفة، باستعمال موضوعات كتبت بلغة مختلفة. كذلك كان المشاركون يجرون أشواطا أحادية اللغة في اللغة المستهدفة لتكون بمثابة الحد الأساس أو العتبة. وكان هذا المسار يركز، في المؤتمر العاشر، على استرجاع الوثائق باستعمال موضوعات بالإنجليزية أو الفرنسية. وقد تم تطبيق هذا المسار لآخر مرة في مؤتمر ٢٠٠٣، إلا أن المزيد من جهود البحث في هذا الموضوع تواصلت في كلف **CLEF** (منتدى التقييم عبر اللغات **Cross Language Evaluation Forum**)، والمجموعة التجريبية لنظم استرجاع المعلومات بالمعهد الوطني للمعلوماتية **National Institute of Informatics' Test Collection for IR Systems (NTCIR)**. وقد تناولنا كلا الموقعين في ٩/ ١/ ٣.

- مسار إدماج مراصد البيانات **Database Merging track**، وقد طبق هذا المسار في المؤتمرين الرابع والخامس فقط، ودرس الطرق الخاصة بمعالجة المجموعات غير المتجانسة، كالخروج بترتيب طبقي واحد للوثائق الخاصة بالاستفسارات، حينما تكون مجموعات البيانات التي يتم الاعتماد عليها مشتتة.

- مسار المشروع **Enterprise track**، الذي بدأ في مؤتمر ٢٠٠٥، بهدف دراسة البحث عن المشروعات (عندما يكون المستفيد بصدد البحث عن بيانات إحدى المنظمات، أو في بيئة مؤسساتية)، وتوقف في مؤتمر ٢٠٠٨.

- مسار الكيانات **Entity track**، وهو أحد مسارين جديدين بدآ في مؤتمر ٢٠٠٩، وكان هدفه العام إجراء عمليات البحث المتصلة بالكيانات في

بيانات العنكبوتية. وتتعلق مهام البحث هذه (كالعثور على الكيانات وخصائص الكيانات على سبيل المثال) بالاحتياجات المعلوماتية المألوفة، التي لم تصبح بعد على غرار عمليات البحث الخاصة عن الوثائق.

• مسار التنقية **Filtering track**، الذي ينظر إليه بوصفه الشكل "الحقيقي" لمهمة التوجيه، التي أدمجت مع هذا المسار منذ المؤتمر السابع. ويتعلق مسار التنقية بالمهام التي تتسم فيها الموضوعات بالثبات، فضلا عن معرفة بعض الوثائق الصالحة، إلا أن هناك تدفقاً في الوثائق الجديدة التي تضاف إلى النظام. ويتعين على النظام، بالنسبة لكل وثيقة، أن يتخذ قرارا ثنائيا، إما استرجاع الوثيقة أو تكوين قائمة طبقية. وقد توقف هذا المسار في مؤتمر ٢٠٠٣، بعد ثمانية أشواط.

• مسار دراسات الجينوم **Genomics track**، وكان مصمما لاستكشاف مهام الاسترجاع، في مجال يبدي اهتماما ببيانات دراسات الجينوم. وقد استمر من مؤتمر ٢٠٠٣ حتى مؤتمر ٢٠٠٧. وقد حددت معالم بيانات دراسات الجينوم في هذا المسار، بصورة عامة، بحيث لا تشمل مجرد سلاسل الجينات فحسب، وإنما لدعم التوثيق أيضا، كالأوراق البحثية وتقارير المختبرات، التي تغطي مدى واسعا من المعلومات في المجال.

• مسار الاسترجاع عالي الدقة من الوثائق **High Accuracy Retrieval track from Documents (HARD**هارد**)**، الذي نشأ لتحقيق الاسترجاع عالي الدقة، بالاستفادة من المعلومات الإضافية حول من يقوم بإجراء عمليات البحث، أو حول سياق البحث. وتشمل الأساليب التي طبقت في هذا المسار استرجاع المقاطع، و التفاعل المستهدف مع من يقوم

بإجراء عمليات البحث. وقد استمر هذا المسار من مؤتمر ٢٠٠٣ حتى مؤتمر ٢٠٠٥.

- مسار التحقيق المرتفع **High Precision track**، الذي اختبر فاعلية وكفاءة واجهات المستفيدين الخاصة بالنظم المشاركة في المؤتمرين السادس والسابع. وباستخدام مجموعة الخمسين نفسها موضوعا، وكما استخدمت الوثاق في المهمة الخاصة، كان يطلب من المستفيد العثور في غضون خمس دقائق، على عشر وثائق تجيب عن السؤال. ولم يكن بإمكان المستفيدين التعاون معا في سؤال واحد، كما لم يكن بإمكان النظام أو المستفيد أن تكون لديه معرفة مسبقة حول السؤال. ومن ناحية أخرى كانت لدى المستفيد حرية الرجوع إلى أي مصدر متاح، مادام زمن الاسترجاع لا يزيد عن الدقائق الخمس.

- المسار التفاعلي **Interactive track**، وهو مهمة خاصة بدراسة تفاعل المستفيد مع نظم استرجاع النصوص، وتطوير مناهج لتقييم التفاعل. وكانت هذه المهمة، في المؤتمر العاشر، تنطوي على ملاحظة الأفراد الذين يفيدون من العنكبوتية الحية، لإنجاز مهام محددة **(Voorhees & Harman, 2002)**. وقد طبق هذا المسار لآخر مرة في مؤتمر ٢٠٠٢، بعد أن أصبح مسارًا في ثمانية مؤتمرات متتابعة.

- المسار القانوني **Legal track**، الذي بدأ في مؤتمر ٢٠٠٦، ليحاول تطوير تقنية بحث تلبي احتياجات المحامين، ومساعدتهم في العثور على المعلومات بفاعلية، في مجموعات الوثائق الرقمية. وكان لا يزال جاريا في مؤتمر ٢٠٠٩.

- مسار المليون استفسار **Million Query track**، الذي بدأ عام ٢٠٠٧، وهو مماثل للاسترجاع الخاص المعياري، ولكن مع اختلاف في التركيز. وقد قصد بهذا المسار على وجه التحديد، تقييم أعداد كبيرة من الاستفسارات (ومن هنا

سمي بالمليون استفسار)، على نحو غير مكتمل، بدلا من عدد قليل من الاستفسارات على نحو أكثر اكتمالا، كما كان يتم في المهام الخاصة السابقة. وظل مسارا في مؤتمر ٢٠٠٩.

- المسار متعدد اللغات **Multilingual track**، الذي كان ينظر في قضايا الاسترجاع في اللغات غير الإنجليزية. وقد أجري اختبار إسباني تمهيدي في المؤتمر الرابع، ليصبح مسارا رسميا في المؤتمر الخامس. كذلك تمت دراسة الاسترجاع بالصينية في المؤتمرين الخامس والسادس. وقد تواصل المزيد من البحث حول استرجاع المعلومات متعددة اللغات، في منتدى التقييم عبر اللغات (كلف **CLEF**) والمجموعة التجريبية لنظم استرجاع المعلومات بالمعهد الوطني للمعلوماتية (**NTCIR**) اللذين تناولناهما في ٩/ ١/ ٣.

- مسار معالجة اللغة الطبيعية **Natural Language Processing (NLP) track**، الذي استكشف، في المؤتمرين الخامس والسادس، ما إذا كانت أساليب معالجة اللغة الطبيعية (النظمية **syntactic** والدلالية **semantic**) المتاحة، ناضجة بما فيه الكفاية، ليكون لها تأثيرها على استرجاع المعلومات، وما إذا كان بإمكان هذه الأساليب أن تحقق تميزا على طرق الاسترجاع الكمية البحت (كالإحصائية مثلا).

- مسار الجدة **Novelty track**، الذي درس، من مؤتمر ٢٠٠٢ حتى مؤتمر ٢٠٠٤، قدرة النظم على العثور على المعلومات الجديدة (أي غير المكررة) والصالحة. وقد اعتمد هذا المسار على خمسين من أسئلة المؤتمرات القديمة لاسترجاع النصوص، باستخدام قائمة طبقية ثابتة للوثائق الصالحة المعروفة.

- مسار الاستفسار **Query track**، الذي بدأ في المؤتمر السابع، لتشجيع البحث حول تأثير قابلية الاستفسارات للتغير، وتحليل الاستفسارات، على

أداء الاسترجاع. وقد قام كل فرد من المشاركين بصياغة عدة أشكال مختلفة، من أسئلة مؤتمرات استرجاع النصوص، الموجودة فعلا، مع التركيز، في معظم الأحيان، على الأشكال المختلفة للأسئلة بلغات طبيعية مختلفة. ثم قامت جميع المجموعات بتطبيق مختلف الصيغ. وقد تحول المسار إلى "محطة للبحوث" بعد المؤتمر التاسع، وموقع لمؤتمرات استرجاع النصوص على العنكبوتية TREC website (trec.nist.gov)، ليكون الآن بمثابة مستودع لكل من صيغ الاستفسارات، ونتائج الاسترجاع التي أسفرت عنها تلك الاستفسارات.

- مسار الإجابة عن الأسئلة **Question Answering (QA) track**، الذي بدأ في المؤتمر الثامن، لاتخاذ خطوة أكثر اقترابا من استرجاع المعلومات (أي استرجاع المقاطع) وليس استرجاع الوثائق. ولأغراض أسئلة الاختبار، أنتجت النظم مقتطفات نصية تجيب على الأسئلة. وكان للأشواط المختلفة حدود مختلفة لأقصى طول للمقتطف، يتراوح بين عبارة موجزة (كلمتان أو ثلاث) إلى الوثيقة برمتها (ألف كلمة). وقد أجرى المؤتمر العاشر دراسة استرشادية، بأسئلة كانت تتطلب معلومات من وثائق متعددة، لتتجمع لتشكل إجابة صحيحة (**Voorhees & Harman, 2002**). وقد توقف هذا المسار بعد مؤتمر ٢٠٠٧.

- مسار تلقيم الصلاحية المرتد **Relevance Feedback Track**، الذي يخطط لوضع إطار لاستكشاف تأثير العوامل المختلفة على نجاح تلقيم الصلاحية المرتد، وهو أسلوب يستخدم على نحو متزايد في استرجاع المعلومات. وقد طبق هذا المسار في مؤتمر ٢٠٠٨، وتواصل في مؤتمر ٢٠٠٩.

• مسار الاسترجاع القوي **Robust Retrieval track**، وهو استمرار لمسار العنكبوتية **Web track**، انطلق في مؤتمر ٢٠٠٣، وكان يشبه مهمة الاسترجاع الخاصة التقليدية، إلا أنه كان يركز أكثر على فاعلية السؤال الواحد، وعلى نحو أقل، على متوسط الفاعلية. وعلى الرغم من توقف هذا المسار بعد مؤتمر ٢٠٠٥، فإن بعض البيانات التي أسفر عنها تستعمل في مهام خاصة أخرى.

• مسار البريد المزعج **SPAM track**، الذي بدأ في مؤتمر ٢٠٠٥، لإجراء تقييم معياري للطرق الحالية والمقترحة لتقنية البريد من الرسائل المزعجة. كما كان من المتوقع أيضا وضع الأساس فيما بعد، لتقييم المزيد من المهام العامة لتنقية البريد الإلكتروني واسترجاعه. وقد توقف هذا المسار بعد مؤتمر ٢٠٠٧.

• مسار استرجاع الوثائق الناطقة **Spoken Document Retrieval (SDR) track**، الذي ورث مسار التشويش، ودرس قدرة نظام الاسترجاع على استرجاع الوثائق الناطقة (أي التسجيلات والأحاديث). وقد بدأ هذا المسار بدعم من جماعة الحديث في المعهد الوطني للمواصفات المعيارية والتقنية (نست **NIST**). وقارن المشاركون فيه فاعلية نظامهم للاسترجاع اعتمادا على نصوص بشرية للأخبار الإذاعية، ونصوص من إنتاج نظام أساس للتعرف على الحديث، واختياريا على نصوص من إنتاج من يتعرفون على الحديث. وقد توقف هذا المسار في المؤتمر العاشر.

• مسار التيرابايت **Terabyte track**، الذي طبق لأول مرة في مؤتمر ٢٠٠٤، وكان هدفه دراسة ما إذا كان من الممكن، وكيف لأوساط استرجاع المعلومات مقارنة التقييم المعتمد على مجموعات تجريبية تقليدية لاسترجاع

المعلومات، ومقابل مراصد بيانات أكبر على نحو ملحوظ من تلك المستخدمة في مؤتمرات استرجاع النصوص وقتئذ. وقد طبق هذا المسار لآخر مرة في مؤتمر ٢٠٠٦.

- مسار الرصيد بالغ الضخامة **Very Large Corpus (VLC) track**، الذي طبق تجريبيا في المؤتمر الخامس، وتوقف بعد المؤتمر السابع، وأنجز عمليات بحث خاصة على حوالي ٢٠ جيجا بايت **20 GB** من النصوص، لاستكشاف فاعلية وكفاءة الاسترجاع. وكان المشاركون فيه يتم تقييمهم بناء على دقة استرجاع الوثائق العشرين المتربعة على القمة، وزمن الاستجابة بالنسبة للاستفسار، وزمن بناء نظام البيانات، فضلاً عن بعض مقاييس التكلفة.

- مسار الفيديو **Video track**، الذي استجد في المؤتمر العاشر، وكان يركز على التعامل القائم على المحتوى مع الفيديو الرقمي، لمواصلة اهتمام مؤتمرات استرجاع النصوص باسترجاع الوسائط المتعددة **(Voorhees & Harman, 2002)**. وكان الاهتمام بهذا المسار واضحاً، على الرغم من وجود تداخل في التقنيات الخاصة باسترجاع الفيديو وتقنيات استرجاع النصوص. ونتيجة لذلك انفصل المسار عام ٢٠٠٣، كجهد تقييم مستقل، يسمى تركفيد **TRECvid (www.nlpir.nest.gov/ projects/trecvid)**.

- مسار العنكبوتية **Web track**، ويمثل مهام خاصة، مجموعة وثائقها مقتطفة من العنكبوتية العالمية **World Wide Web**. وتنطوي المهمة الرئيسة على حوالي ١٠ جيجا بايت **10 GB** من بيانات العنكبوتية. وكان المشاركون فيه يحظون بالتشجيع لدراسة مجموعة متنوعة من استراتيجيات استرجاع العنكبوتية، على غرار ما إذا كان من الممكن استخدام الروابط لزيادة سرعة

الاسترجاع. وقد بدأ هذا المسار في المؤتمر الثامن، وتوقف بعد مؤتمر ٢٠٠٤. إلا أن مسار العنكبوتية أعيد ثانية في مؤتمر ٢٠٠٩ لاستكشاف مهام استرجاع ترتبط بالعنكبوتية على وجه التحديد، تشمل مهام التنوع والكفاءة، ومجموعات تتكون من ما يصل إلى بليون صفحة من صفحات العنكبوتية. وبالمقارنة بسلفه فيما بين المؤتمر الثامن ومؤتمر ٢٠٠٤، كان من الممكن للمسار الراهن أن يستخدم زواحف عنكبوتية **web crawl** أكبر وأحدث على غرار مجموعة الوثائق.

ويوضح وصف جميع المسارات هذا، حجم ومدى مهام الاسترجاع التي تمت دراستها في مؤتمرات استرجاع النصوص. وقد تجاوزت بعض المسارات فعلا الأهداف التي حددت لها في البداية؛ فمؤتمرات استرجاع النصوص في جوهرها، على سبيل المثال، مشروع لتقييم استرجاع النصوص، إلا أن مسارًا حول استرجاع الفيديو طبق فعلا في المؤتمر العاشر. ومثل هذه المسارات، كما ذهبت سبارك جونز **Spärck Jones (2000)**، امتدادات لمؤتمرات استرجاع النصوص. يضاف إلى ذلك، كما ألمح هارمان وفورهيس **Harman and Voorhees (2006)**، في الشكل رقم ٤/ ٣ في بحثهما، إلى أن معظم المسارات التي طبقت بدءا من المؤتمر الأول حتى مؤتمر ٢٠٠٣، تدخل في الفئات التالية (مع إضافة المسارات التي بدأت بعد مؤتمر ٢٠٠٣، حيثما كان ذلك ملائما):

- النصوص الثابتة: المسارات، والخاص، والقوي، والتلقيم المرتد للصلاحية.
- النصوص المتدفقة: التوجيه، والتنقية.
- البشرية: التفاعلي والاسترجاع عالي الدقة من الوثائق **HARD**

• خارج نطاق الإنجليزية: متعدد اللغات، واسترجاع المعلومات عبر اللغات **CLIR**.

• خارج نطاق النصوص: التشويش، واسترجاع الوثائق الناطقة، والفيديو، وتركفيد **TRECVid**.

• البحث في العنكبوتية: الرصيد بالغ الضخامة، والعنكبوتية، والتيرابايت، والكيانات.

• الإجابات وليست الوثائق: إجابة الأسئلة، والجدة.

• المجال: دراسات الجينوم، والمشروع، والقانوني، واسترجاع المعلومات الكيميائية.

التقييم وأحكام الصلاحية:

استخدم التقييم في مؤتمرات استرجاع النصوص **TREC**، مقياسي الاستدعاء والتحقيق التقليديين، ومن ثم فقد كانت أحكام الصلاحية ضرورة لا فكاك منها. وكان أولئك الذين صاغوا الأسئلة، مشاركين أيضا في التقييم أحيانا. وكانت طريقة التجميع البؤري **pooling**، التي تناولناها في ١١/ ١/ ٢/ ١/ ٢، تستخدم لتشكيل تجمع البيانات للتقييم، في جميع مؤتمرات استرجاع النصوص. وكان كل فريق مشارك يقدم للرعاة الوثائق المئة المسترجعة المتربعة على القمة، في كل شوط، وذلك لسؤال بعينه؛ نظرا لتطبيق خوارزمية للترتيب الطبقي في جميع نظم الاختبار، وذلك على نتائج مخرجات الاسترجاع، بناء على الصلاحية المنتظرة.

وكان كل مشارك في التقدير يحكم في المتوسط على حوالي ٢٠٠٠ وثيقة، لكل سؤال. وكان اكتمال الأحكام واطرادها في مؤتمرات استرجاع النصوص، يراجعان في دراسات منفصلة، وقد بدت النتائج تبدو إيجابية **(Voorhees & Harman, 2000)**.

وكانت هناك حزمة تقييم معيارية، تستخدم لتسجيل نتائج كل شوط من الأشواط التي يتم إنجازها. وكان مقياس التقييم الأكثر كثافة في الاستعمال، في مؤتمرات استرجاع النصوص، هو التحقيق المتوسط الذي نوقش في ١١/ ١/ ٢/ ١/ ٤. كما كان يتم أيضاً حساب أنواع أخرى من مقاييس التقييم المعتمدة على الاستدعاء والتحقيق، في مؤتمرات استرجاع النصوص. وقد لخصت سبارك جونز **Spärck Jones (2000)** مقاييس أداء المسارات التي طبقت، بدءا من المؤتمر الرابع حتى المؤتمر السادس. كذلك قدم باكلي وفورهس **Buckley and Voorhess (2005)** وصفا ضافياً لمقاييس التقييم التي استخدمت في مؤتمرات استرجاع النصوص. إلا أن جميع مقاييس التقييم لم تكن تقيم سوى أداء الاسترجاع؛ نظراً لأنها كانت جميعها أشكالاً مختلفة للتحقيق والاستدعاء، تاركة الأبعاد الأخرى لنظم استرجاع المعلومات دون مساس. ومن ثم، فإن سلسلة مؤتمرات استرجاع النصوص، تعالج مقاييس التحقيق والاستدعاء المرتفعين، بوصفهما هدفها النهائي، وليسا هدفا فرعيا **(Robertson, 2008)**.

نتائج مؤتمرات استرجاع النصوص:

لقد أسفرت مؤتمرات استرجاع النصوص عن نتائج كثيرة جداً، متنوعة بحيث يصعب حصرها في هذا السياق؛ ومن ثم فإننا نلخص النتائج العامة فقط على النحو التالي:

- هناك كثير من الطرق المختلفة تمام الاختلاف التي تؤدي على مستوى مشابه، الأمر الذي يدل على أن "جميع الطرق تفضي إلى روما"، فمن الممكن للطرق المختلفة أن تسعى لتحقيق الهدف نفسه.
- يعمل التجهيز الآلي (أي صياغة الاستفسارات أو ناتج الاسترجاع) بكفاءة التجهيز اليدوي نفسها إلا أنه قد لا يبدو مقنعاً تماماً، بالنسبة للاستفسارات البسيطة.

- تبدو أساليب التكشيف والبحث، القائمة على الإحصاء، منخفضة التكلفة وتنافسية، وخصوصاً بالنسبة للمجموعات أو مراصد البيانات الضخمة. ومن الممكن لهذه الأساليب أن تعمل على نحو أفضل بكثير، من الخطوط الأساس للمصطلحات البسيطة، على الرغم من أنه من الممكن لهذه الأساليب أن تصنف ضمن نوع ما من التدريب الجماعي.

- تسهم العبارات البسيطة، إلى جانب المصطلحات المفردة، بشيء ما، مما يدل على وعد إضافي بالنسبة للطرق الآلية في المجال.

- لا يبدو توافر تلقيم الصلاحية المرتد، وكذلك التوسع المعتدل في الاستفسارات مفيداً. إلا أنه ليس هناك في مؤتمرات استرجاع النصوص، تقدم يذكر نحو هدف صياغة استفسار واحد فعال، وذلك في مقابل تصنيف الاستفسارات والاستراتيجيات وفقاً للمهام الخاصة. وبعبارة أخرى، فإن النظام لا يزال عاجزاً عن تحليل الطلب، وفقاً لمضامين صياغة مصطلحاته آلياً، واختيار استراتيجية معينة لصياغة الاستفسارات، يمكن أن تسفر عن أفضل النتائج.

مشكلات مؤتمرات استرجاع النصوص:

النقد الرئيس الذي يوجه لمؤتمرات استرجاع النصوص، أنها سارت في الأساس على خطى تجارب كرانفيلد، باستعمال بيئة اختبار اصطناعية، وباتخاذ الاستدعاء والتحقيق مقياسين لأداء الاسترجاع.

واصطناعية سلسلة مؤتمرات استرجاع النصوص ثلاثية الأبعاد؛ أولها أن الوثائق التجريبية التي استخدمت في المؤتمرات المبكرة، التي كانت تتكون في المقام الأول من الصحف، كانت تجمع خصوصاً للمشروع، ولم تكن مجموعة تكونت طبيعياً

على مر السنين. وبينما يمكن للصحف أن تكون مصدراً قيماً بالنسبة لأغراض أخرى كثيرة، فإنها لا يرجع إليها بكثافة فيما يتعلق بمعظم جهود البحث العلمي. ثانياً، لم يكن لسلسلة مؤتمرات استرجاع النصوص مستفيدون حقيقيون متمرسون؛ فقد كانت الأسئلة أو طلبات البحث، كما في كرانفيلد، مصنعة خصوصاً من أجل المشروع. وكان الاسترجاع يتم بواسطة من قاموا بتطوير النظام، كما كان أداء الاسترجاع يحكم عليه من قبل مقيمين ربما يكونون ممن صاغوا الأسئلة أيضاً. ثالثاً، كانت التجارب تتم في بيئة مختبر تقليدي، بدرجة معينة من الانضباط والمعالجة، ولا يمكن ربط مثل هذه البيئة التجريبية بالنظم العاملة الحقيقية.

وفضلاً عن ذلك، هناك مشكلات تتعلق بطريقة التقييم المتبعة في مؤتمرات استرجاع النصوص؛ فأوجه القصور التي تكتنف استخدام الاستدعاء والتحقيق كمقياسين لتقييم أداء الاسترجاع، من ناحية، ماثلة تماماً في مؤتمرات استرجاع النصوص؛ نظراً لأن الاستدعاء و التحقيق هما المقياسان الرئيسان اللذان وقع عليهما الاختيار للتقييم في المشروع. ومن ناحية أخرى، لم يتم لأغراض التقييم، تجميع سوى النتائج المتربعة على القمة، المسترجعة من كل نظام؛ ومن ثم، فإن تقديرات مؤتمرات استرجاع النصوص يمكن أن تبدو أفضل مما كان يمكن أن تكون عليه، في حالة تغطية التقييم لكل النتائج. ووفقاً لبحث أعده زوبل **Zobel (1998)**، فإن الاستدعاء في مؤتمرات استرجاع النصوص، كان تقديره ينطوي على مبالغة. وفضلاً عن ذلك، تبدو الحاجة إلى المزيد من التقييم الإحصائي واضحة؛ نظراً لأن منحنيات الاستدعاء في مقابل التحقيق، التي قدمها كل فريق مشارك، كانت متقاربة في معظم الأحيان، وهناك حاجة للتحقق مما إذا كانت هناك فعلاً أي فروق ذات دلالة إحصائية بين نتائج نظامين، أو مجموعتين من نتائج النظام نفسه. كذلك يبدو من الضروري تجاوز حدود المتوسطات؛ لفهم أداء النظم على نحو أفضل. يضاف إلى ذلك أنه من الصعب بمكان

إجراء تحليل للإخفاق على النتائج؛ نظراً لضخامة عدد الوثائق التجريبية التي شملتها التجارب، والحد الموضوع لمخرجات الاسترجاع (Harman, 1995). وقد تمت الإشارة أيضاً إلى غياب تحليل الإخفاق في سلسلة مؤتمرات استرجاع النصوص، من جانب روبرتصون **Robertson (2008)**.

وتتبع مؤتمرات استرجاع النصوص إجمالاً، نهج الصندوق الأسود، بالنسبة للتقييم؛ إذ تتم المقارنة بين النظم ككتل لأغراض قياس الأداء، عن طريق النتائج التي استرجعتها فرق البحث (Spärck Jones, 2000). ولم يتم التوصل إلى معلومات تذكر حول ما يمكن أن يكون لاستراتيجيه معينة، أو أداة من قدرة على تغيير أداء النظام، وذلك على الرغم من أن مؤتمرات استرجاع النصوص كانت تعقد منذ عام ١٩٩٢. وهناك حاجة واضحة إلى نهج الصندوق الزجاجي الشفاف من أجل مهمة التقييم. وكررت سبارك جونز **Spärck Jones (2006)** الإعراب عن هذه الحاجة، بالدعوة إلى انتباه مؤتمرات استرجاع النصوص للتعامل لأدق الدقائق المميزة، في تقييم استرجاع المعلومات.

أهمية مؤتمرات استرجاع النصوص:

بكثير من خصائصها الفريدة (كحجم الوثائق التجريبية والعدد المتزايد للفرق المشاركة على سبيل المثال) تشغل مؤتمرات استرجاع النصوص موقعاً مهماً في تاريخ تقييم تنظيم المعلومات واسترجاعها. وقد اتضحت أهمية سلسلة المؤتمرات هذه في الجوانب التالية:

أولاً، أنها توفر منتدى منفتحاً ومشجعاً بالنسبة لتقييم تنظيم المعلومات واسترجاعها. ولأنها منتدى منفتح؛ فقد شجعت على المشاركة من جانب كثير من المؤسسات والمنظمات في الداخل والخارج. ولأنها منتدى مشجع؛ فقد حثت المشاركين

على إدخال المزيد من التحسينات على نظمهم، فقد حدث بالفعل على سبيل المثال، أن تمت ملاحظة تحسينات مهمة في أداء الاسترجاع، في المؤتمر الثاني عما كان عليه الحال في المؤتمر الأول (**Spärck Jones, 1995**).

ثانياً، تمت تجربة الكثير من الأساليب المختلفة في مؤتمرات استرجاع النصوص، بما في ذلك استرجاع المقاطع أو الفقرات، ودمج البيانات، والتكشيف الإحصائي والتكشيف الاحتمالي، واستراتيجيات وزن المصطلحات. أما تقنية الاسترجاع البولينية، فعلى الرغم من أهميتها، فإنها لم تكن لها الغلبة في المشروع. وفضلاً عن ذلك، فإن حجم المجموعات التجريبية يبلغ الآن ما يقدر بالجيجابايت لا بالميجابايت، كما كان في جهود التقييم المبكرة، وبذلك تصبح هذه المجموعات أكثر اقتراباً من موقف الحياة الواقعية، من حيث الحجم. ولقد كانت مؤتمرات استرجاع النصوص تتطور نحو ما يضاهي بيئة العمل الواقعية، التي تكفل نقل نتائج التجارب لتدخل في المنتجات التجارية، وذلك بعائدات لا يمكن إنكارها (**Kowalski & Maybury, 2000**). ونظم استرجاع الإنترنت أمثلة بارزة لثمار بحوث مؤتمرات استرجاع النصوص. ولقد كانت قدرة هذه النظم على توجيه المستفيدين نحو ما يبحثون عنه من معلومات، أمراً جوهرياً بالنسبة لنجاح العنكبوتية .(**Harman & Voorhees, 2006**)

ثالثاً، تبدو مؤتمرات استرجاع النصوص أكثر قابلية للتعميم، من كثير من المشروعات الأخرى، ويرجع ذلك جزئيا، إلى أن هذه المؤتمرات جهد بحثي متواصل. إذ إنها تحاول الإجابة على كثير من التساؤلات، التي يمكن وضعها في مثل هذه الفئات العريضة، كنماذج التكشيف والاسترجاع، ولغات التكشيف، ووصف الوثائق، ومصادر التكشيف، والاستفسارات، ومصادر الاستفسارات، واستراتيجيات البحث، ومعايير احتساب النقاط، وقدرة نظم استرجاع المعلومات على التعلم .(**Spärck Jones, 2000**)

ويمكن القول إجمالاً، إن مؤتمرات استرجاع النصوص على قدر كبير من الأهمية بالنسبة للمجال. كذلك تم اقتراح احتمالات بحوث جديدة للمستقبل، تشمل:

- عقد مؤتمرات استرجاع النصوص، على فترات متباعدة، بحيث يتاح الوقت لهضم ما يتم الحصول عليه فعلاً من نتائج.
- تركيز الجهد على مهام أو مسارات أقل، من أجل مواصلتها بمزيد من التعمق.
- إدخال مستفيدين حقيقيين في المشروع، للارتفاع بمستوى قابلية نتائج البحوث للتطبيق.
- وضع تطبيق نتائج تجارب استرجاع المعلومات في الصدارة، ودفع مؤتمرات استرجاع النصوص على نحو أكثر حزماً، تجاه دراسات الاسترجاع في سياقات عملية.
- اتباع نهج الصندوق الزجاجي الشفاف في التقييم، بإشراك الفرق كل على حدة، في الاستراتيجية أو الأداة المحددة، وربما أيضاً على مستوى العامل المحدد، بدلا من اتباع نهج الصندوق الأسود، حيث تقارن النظم إجمالاً من حيث الأداء، عن طريق النتائج التي يحصل عليها كل فريق.
- الشروع في إجراء دراسات متعددة المهام، بتطبيق خبرة مؤتمرات استرجاع النصوص في هذه الدراسات. ومن شأن الاسترجاع أن تتم دراسته بوصفه أحد مكونات النظام، الذي يمكن أن يكفل أيضاً التلخيص الآلي للنصوص المسترجعة، على سبيل المثال.

كلمة أخيرة حول تقييم تنظيم المعلومات واسترجاعها:

ترمز تجارب كرانفيلد، وسلسلة مؤتمرات استرجاع النصوص، للمشكلات والقضايا التي ينطوي عليها تقييم تنظيم المعلومات واسترجاعها. ويمكن القول بوجه عام، أن معظم مشروعات التقييم يتم تنفيذها في المختبرات، لا في مواقف الحياة الحقيقية، وذلك بمجموعات اصطناعية، وطلبات مصنعة. وعادة ما يكون حجم المجموعة التجريبية صغيراً، وإن كانت سلسلة مؤتمرات استرجاع النصوص تعد استثناءً. ويستخدم الاستدعاء والتحقيق، والتسرب **fallout**، كمقاييس تقييم في غالب الأحيان. إلا أنه لا يمكن التعويل عليها؛ لأنها من بين المخاوف الأخرى، تقوم على أحكام ذاتية للصلاحية. وفضلاً عن ذلك، تركز معظم مشروعات التقييم على هذه المعايير كل على حدة (كالاستدعاء والتحقيق، على سبيل المثال) في غياب إطار شامل لتقييم النظام برمته. والخلاصة: أننا نعلم ما يحدث في مثل هذه النظم، إلا أنه لا يتضح لنا تماماً، ما يتعلق بأسباب حدوث ذلك، وكيف يحدث، نظراً لنهج الصندوق الأسود للتقييم.

وقد طرح الكثير من الأفكار والمقترحات، لدفع البحث في تقييم تنظيم المعلومات واسترجاعها قدما، بالخروج عن حدود النماذج القائمة، وتطبيق التعددية المنهجية؛ فقد وصف هارتر وهيرت **Harter and Hert (1997)**، على سبيل المثال، إطاراً ناشئاً لتقييم تنظيم المعلومات واسترجاعها، يركز على استخدام الأبعاد والطرق المتعددة، والتقييم من منظور عدة مشاركين في أنشطة تنظيم المعلومات واسترجاعها. كما أنهما يؤمنان بقوة، بأن تفاعل المستفيد مع النظام ينبغي أن يوضع في الحسبان، أثناء عملية التقييم، في العصر الرقمي. ومن ناحية أخرى، فإن مشروعات التقييم المستقبلية ينبغي أن تتم في بيئة واقعية، وتستخدم مقاييس أخرى، خلاف الاستدعاء والتحقيق والتسرب، وذلك للتغلب على وجهي القصور الأسوأ تأثيراً في جهود التقييم السابقة في المجال؛ وهما الذاتية في الحكم على الصلاحية، وعدم القدرة على تحديد العدد الإجمالي للوثائق الصالحة في النظام.

المراجع

Antelman, Kristin, Lynema, Emily, and Pace, Andrew K. (2006). Toward a twenty-first century library catalog. *Information Technology & Libraries,* 25(3), 128-139.

Bell, Hazel K. (1991). Bias in indexing and loaded language. *The Indexer,* 17(3), 173-177.

Bevan, Nigel. (2001). International standards for HCI and usability. *International Journal of Human Computer Studies,* 55(4), 533-552.

Blair, D. C., and Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system. *Communications of the ACM,* 28(3), 289-299.

Buckley, Chris, and Voorhees, Ellen M. (2005). Retrieval system evaluation. In Ellen M. Voorhees and Donna K. Harman (Eds.), *TREC: Experiment and evaluation in information retrieval* (pp. 53-75). Cambridge, MA: MIT Press.

Chang, Shih-Fu. (2002). The holy grail of content-based media analysis. *IEEE Multimedia Magazine,* 9(2), 6-10.

Chowdhury, Gobinda G. (1999). *Introduction to modern information retrieval.* London: Library Association Publishing.

Chu, Heting, and Rosenthal, Marilyn. (1996). Search engines for the World Wide Web: A comparative study and evaluation methodology. *Proceedings of the 59th Annual Meeting of the American Society for Information Science,* 33, 127-135. Retrieved January 21, 2009, from www.asis.org/ annual-96/ElectronicProceedings/chu.html

Clarke, Sarah J., and Willett, Peter. (1997). Estimating the recall performance of web search engines. *Aslib Proceedings,* 49(7), 184-189.

Cleverdon, C. W. (1962). *Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems.* Cranfield, England: College of Aeronautics.

Cleverdon, C. W. (1967). The Cranfield tests on indexing language devices. *Aslib Proceedings,* 19,173-192.

Cleverdon, C. W. (1984). Optimizing convenient online access to bibliographic databases. *Information Services and Use,* 4(1-2), 37-47.

Cleverdon, C. W., and Mills, J. (1963). The testing of indexing language devices. *Aslib Proceedings,* 15(4), 106-130.

Cleverdon, C. W., Mills, J., and Keen, E. M. (1966). *Factors determining the performance of indexing systems, Vol. 1-Design.* Cranfield, England: Aslib Cranfield Research Project.

Clough, Paul. (2007). Large-scale evaluation of cross-language image retrieval systems. *Bulletin of the American Society for Information Science and Technology,* 33(3). Retrieved December 19, 2008, from www.asis.org/ Bulletin/Feb-07/clough.html

Cooper, William S. (1973a). On selecting a measure of retrieval effectiveness. Part *I. Journal of the American Society for Information Science,* 24, 87-100.

Cooper, William S. (1973b). On selecting a measure of retrieval effectiveness. Part II. *Journal of the American Society for Information Science,* 24, 413-124.

Deselaers, Thomas, Keysers, Daniel, and Ney, Hermann. (2008). Features for image retrieval: An experimental comparison. *Information Retrieval,* 11(2), 77-107.

Ding, Wei, and Marchionini, Gary. (1996). A comparative study of web search service performance. *Proceedings of the 59th Annual Meeting of the American Society for Information Science,* 33, 136-142.

Downie, J. Stephen. (2003). The TREC-like evaluation of music retrieval systems. *Proceedings of the 26th annual international*

ACM SIGIR conference (pp. 453-454). New York: Association for Computing Machinery.

Downie, J. Stephen. (Ed.). (2004). The MIR/MDL Evaluation Project white paper collection. 3rd ed. Retrieved January 7, 2009, from www.music-ir.org/ evaluation/wp.html

Foskett, A. C. (1996). *Subject approach to information.* 5th ed. London: Library Association Publishing.

Froehlich, Thomas J. (1994). Relevance revisited-Towards an agenda for the 21st century: Introduction to special topic issue on relevance research. *Journal of the American Society for Information Science,* 45(3), 123-134.

Fugmann, Robert. (1993). *Subject analysis and indexing: Theoretical foundation and practical advice.* Frankfurt/Main: IndeksVerlag.

Griffith, Belver C. (1986). *Evaluation of information systems: Class notes.* Philadelphia: Drexel University, College of Information Studies.

Griffiths, Alan, Luckhurst, H. Claire, and Willett, Peter. (1986). Using inter-document similarity information in document retrieval systems. *Journal of the American Society for Information Science,* 37(1), 3-11.

Guha, Tamal Kumar, and Saraf, Veena. (2007). OPAC usability: Assessment through verbal protocol. *Electronic Library,* 23(4), 463-473.

Haga, Hirohide, and Kaneda, Shigeo. (2005). A usability survey of a contents-based video retrieval system by combining digital video and an electronic bulletin board. *Internet & Higher Education,* 8(3), 251-262.

Harman, Donna. (1993). A special conference report: The first Text Retrieval Conference (TREC-1) Rockville, MD, U.S.A., 4-6 Nov. 1992. *Information Processing & Management,* 29(4), 411-414.

Harman, Donna. (1995). Overview of the second Text Retrieval Conference (TREC-2). *Information Processing & Management,* 31(3), 271-290.

Harman, Donna K., and Voorhees, Ellen M. (2006). TREC: An overview. *Annual Review of Information Science and Technology,* 40,113-155.

Harter, Stephen P. (1971). The Cranfield II relevance assessments: A critical evaluation. *Library Quarterly,* 41, 229-243.

Harter, Stephen P., and Hert, Carol A. (1997). Evaluation of information retrieval systems: Approaches, issues, and methods. *Annual Review of Information Science and Technology,* 32,3-94.

Hearst, Marti, et al. (2002). Finding the flow in Web site search. *Communications of the ACM,* 45(9), 42-49.

Hildreth, Charles R. (1982). *Online public access catalogues: The user interface.* Dublin, OH: OCLC.

Hildreth, Charles R. (1985). Online public access catalogs. *Annual Review of Information Science and Technology,* 20, 233-285.

Hooper, R. S. (1965). *Indexer consistency tests-Origin, measurements, results and utilization.* Bethesda, MD: IBM.

Ingwersen, Peter. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation,* 52(1), 3-50.

International Organization for Standardization (ISO). (1998). *Guidance on usability* (ISO 9241-11). Geneva, Switzerland: ISO.

Keen, E.M. (1971). Evaluation parameters. In Gerard Salton (Ed.), *The SMART retrieval system: Experiments in automatic document processing* (pp. 74-111). Englewood Cliffs, NJ: Prentice Hall.

Kent, Allen, et al. (1955). Machine literature searching. VIII. Operational criteria for designing information retrieval systems. *American Documentation,* 6(2), 93-101.

Korfhage, Robert R. (1997). *Information storage and retrieval.* New York: John Wiley & Sons.

Kowalski, Gerald, and Maybury, Mark T. (2000). *Information storage and*
retrieval systems: Theory and implementation. 2nd ed. Boston: Kluwer
Academic Publishers.

Lancaster, F. W. (1968). *Evaluation of the MEDLARS demand search service.* Bethesda, MD: National Library of Medicine.

Lancaster, F. W. (1979). *Information retrieval systems: Characteristics, testing and evaluation.* 2nd ed. New York: Wiley.

Lancaster, F. W. (1998). *Indexing and abstracting in theory and practice.* 2nd ed. Champaign-Urbana, IL: University of Illinois, Graduate School of Library and Information Science.

Lancaster, F. W., and Fayen, E. G. (1973). *Information retrieval online.* Los Angeles: Melville.

Lancaster, F. W., and Warner, Amy J. (1993). *Information retrieval today.* Arlington, VA: Information Resources Press.

Large, Andrew, Tedd, Lucy A., and Hartley, Richard J. (1999). *Information seeking in the online age: Principles and practice.* London: Bowker-Saur.

Leighton, H. Vernon. (1995). Performance of four World Wide Web (WWW) index services: Infoseek, Lycos, WebCrawler, and WWWWorm. Retrieved October 11, 2009, from www.winona.edu/library/staff/vL/WEBIND.HTM

Leighton, H. Vernon, and Srivastava, Jaideep. (1999). First 20 precision among World Wide Web search services (search engines). *Journal of the American Society for Information Science,* 50(10), 870-881.

Lombardo, Shawn V, and Condic, Kristine S. (2000). Empowering users with a new online catalog. *Library Hi Tech,* 18(2), 130-141.

Marcum, Deanna B. (2006). The future of cataloging. *Library Resources & Technical Services,* 50(1), 5-9.

Markey, Karen. (2007). The online library catalog: Paradise lost or paradise gained. *D-Lib Magazine,* 13(1/2). Retrieved December 6, 2008, from dlib.org/dlib/january07/markey/01markey.html

Miller, C., and Tegler, P. (1986). Online searching and the research process. *College and Research Libraries,* 47(4), 370-373.

Miniwatts Marketing Group. (2009). Internet world users by language. Retrieved October 8, 2009, from www.internetworldstats.com/stats7.htm

Mizzaro, Stefno. (1997). Relevance: The whole history. *Journal of the American Society for Information Science,* 48(9), 801-832.

Nielsen, Jakob. (1993). *Usability engineering.* New York: Morgan Kaufmann.

Nielsen, Jakob. (2003). Usability 101: Introduction to usability. Retrieved January 8,2009, from www.useit.com/alertbox/20030825.html

Oppenheim, Charles, et al. (2000). The evaluation of WWW search engines. *Journal of Documentation,* 56(2), 190-211.

O'Rourke, Victoria. (1987). Selection of an online public access catalog: A checklist approach. *Information Technology and Libraries,* 6(4), 278-287.

Petrelli, Daniela, and Auld, Daniel. (2008). An examination of automatic video retrieval technology on access to the contents of an historical video archive. *Program: Electronic Library & Information Systems,* 42(2), 115-136.

Robertson, Stephen E. (1969). The parametric description of retrieval tests. Part I: The basic parameters. *Journal of Documentation,* 25(12), 1-27.

Robertson, Stephen E. (2008). On the history of evaluation in IR. *Journal of Information Science,* 34(4), 439-456.

Rowley, Jennifer E., and Hartley, Richard J. (2008). *Organizing knowledge: An introduction to managing access to information.* Burlington, VT: Ashgate.

Salmon, Stephen R. (1983). Characteristics of online public catalogs. *Library Resources & Technical Services,* 27(1), 36-67.

Salton, Gerard. (1981). The SMART environment for retrieval system evaluation: Advantages and problem areas. In Karen Spärck Jones (Ed.), *Information retrieval experiment* (pp. 316-329). London: Butterworth.

Salton, Gerard. (1992). The state of retrieval system evaluation. *Information Processing & Management,* 28(4), 441-449.

Saracevic, Tefko. (1975). Relevance: A review of the literature and a framework for thinking on the notion in information science. *Journal of the American Society for Information Science,* 26(6), 321-343.

Saracevic, Tefko. (2007a). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology,* 58(13), 1915-1933.

Saracevic, Tefko. (2007b). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology,* 58(13), 2126-2144.

Saracevic, Tefko, et al. (1988). A study of information seeking and retrieving. Part I: Background and methodology. *Journal of the American Society for Information Science,* 39(3), 161-176.

Saracevic, Tefko, and Kantor, Paul. (1988). A study of information seeking and retrieving. Part II: Users, questions, and effectiveness. *Journal of the American Society for Information Science,* 39(3), 177-196.

Schamber, Linda. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology,* 29, 3-48.

Schamber, Linda, Eisenberg, Michael B., and Nilan, Michael S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. *Information Processing & Management,* 26(6), 755-776.

Shneiderman, Ben, and Plaisant, Catherine. (2005). *Designing the user interface: Strategies for effective human-computer interaction.* 4th ed. Boston: Addison-Wesley.

Smeaton, Alan F. (2007). TRECVid: Video evaluation. *Bulletin of the American Society for Information Science and Technology,* 33(3). Retrieved December 19, 2008, from www.asis.org/Bulletin/Feb-07/smeaton.html

Smeaton, Alan F, and Over, Paul. (2003). The TREC-2002 video track report. *Proceedings of the Eleventh Text REtrieval Conference (TREC2002).* Retrieved January 7, 2008, from trec.nist.gov/pubs/trecll/papers/ VIDEO.OVER.pdf

Spärck Jones, Karen. (1981). The Cranfield tests. In Karen Spärck Jones (Ed.), *Information retrieval experiment* (pp. 256-284). London: Butterworths.

Spärck Jones, Karen. (1995). Reflections on TREC. *Information Processing & Management,* 31(3), 191-314.

Spärck Jones, Karen. (2000). Further reflections on TREC. *Information Processing & Management,* 36(1), 37-85.

Spärck Jones, Karen. (2006). What's the value of TREC? Is there a gap to jump or a chasm to bridge? *ACM SIGIR Forum,* 40(1), 10-20.

Su, Louise T. (1991). *An investigation to find appropriate measures for evaluating interactive information retrieval.* Unpublished Ph.D. Dissertation. New Brunswick, NJ: Rutgers, the State University of New Jersey.

Swanson, Don R. (1965). The evidence underlying the Cranfield results. *Library Quarterly,* 35(1), 1-20.

Swanson, Don R. (1971). Some unexplained aspects of the Cranfield tests of indexing performance factors. *Library Quarterly,* 41(3), 223-228.

Swanson, Rowena Weiss. (1978). Performing evaluation studies in information science. In D. W. King (Ed.), *Key papers in design and evaluation of retrieval systems* (pp. 58-74). New York: Knowledge Industry.

Swets, John A. (1963). Information retrieval systems. *Science,* 141, 245-250.

Swets, John A. (1969). Effectiveness of information retrieval methods. *American Documentation,* 20(1), 72-89.

Trant, Jennifer. (2004). Image retrieval benchmark database service: A needs assessment and preliminary development plan. *A Report Prepared for the Council on Library and Information Resources and the Coalition for Networked Information.* Retrieved December 18,2008, from www.clir.org/ pubs/reports/trant04/tranttext.htm

TREC. (2000). Overview [of TREC]. Retrieved June 17, 2001, from trec.nist.gov/ overview.html

Vickery, B. C. (1966). Review of Cleverdon, Mills, and Keen. *Journal of Documentation, 22,* 247-249.

Vickery, B.C. (1967). Review of Cleverdon and Keen. *Journal of Documentation,* 23, 338-340.

Voorhees, Ellen M. (2000). Report on TREC-9. *SIGIR Forum,* 34(2), 1-8.

Voorhees, Ellen M., and Harman, Donna. (2000). Overview of the sixth Text REtrieval Conference (TREC-6). *Information Processing & Management,* 36(1), 3-35.

Voorhees, Ellen M., and Harman, Donna K. (2002). Overview of TREC 2001. *Proceedings of the Tenth Text REtrieval Conference*

(TREC 2001) (pp. 1-15). NIST Special Publication 500-250. Retrieved September 2, 2002, from trec.nist.gov/pubs/treclO/papers/overview_10.pdf

Voorhees, Ellen M., and Harman, Donna K. (Eds.). (2005). *TREC: Experiment and evaluation in information retrieval.* Cambridge, MA: MIT Press.

Wellisch, Hans H. (1995). *Indexing from A to Z.* 2nd ed. New York: H.W. Wilson,

Zobel, J. (1998). How reliable are the results of large-scale information retrieval experiments? In *Proceedings of the 21 Annual International Conference on Research and Development in Information Retrieval- SIGIR-98* (pp. 307-314). New York: Association for Computing Machinery.

الفصل الثاني عشر

الذكاء الاصطناعي في تنظيم المعلومات واسترجاعها

الذكاء الاصطناعي **artificial intelligence (AI)**، مجال يهدف إلى جعل الآلات قادرة على النهوض بمهام، يمكن أن تستلزم ذكاءًا، إذا ما نهض بها البشر. وتبدي أوساط تنظيم المعلومات واسترجاعها اهتماما واضحاً ببحوث الذكاء الاصطناعي، أملا في استثمار تطبيقاته في تنظيم المعلومات واسترجاعها. وعلى الرغم من تآلف الباحثين في الذكاء الاصطناعي، مع الإثارة والإحباط أيضا، في السعي لتحقيق أهدافهم، فإن حماسهم لم يفتر، وإنما على العكس، جرت أعداد متزايدة من المحاولات لتطوير تطبيقات الذكاء الاصطناعي، لأغراض تنظيم المعلومات واسترجاعها، وخاصة منذ أصبحت العنكبوتية العالمية المنصة الرئيسة لتنظيم المعلومات واسترجاعها. ويلقي هذا الفصل نظرة شاملة على بحوث الذكاء الاصطناعي، في سياق تنظيم المعلومات واسترجاعها. ويحظى قطاعان نشطان لبحوث الذكاء الاصطناعي لأغراض تنظيم المعلومات واسترجاعها؛ وهما معالجة اللغة الطبيعية **(NLP)**، والعنكبوتية الدلالية- بالمناقشة، بالإضافة إلى أوجه قصور بحوث الذكاء الاصطناعي واحتمالاته بالنسبة لتنظيم المعلومات واسترجاعها.

نظرة عامة على بحوث الذكاء الاصطناعي:

لقد استقر الذكاء الاصطناعي كمجال للبحث، منذ عام ١٩٥٦ على الأقل، عندما عقد مؤتمر دارتماوث **Spärck & Jones (1991) Dartmouth Conference**. ويهدف الذكاء الاصطناعي لتطوير نظم تتصرف بذكاء، باستخدام

الحاسبات، لإنجاز مهام تستلزم البرهان على أساس المعرفة الواعية. ويمكن أن تفسر المعرفة في الذكاء الاصطناعي، بوصفها تجمعا من المعلومات والحقائق والقواعد المعتمدة؛ أي إن مجرد الحدس والفطرة السليمة **common sense** لا غنى عنهما بالنسبة للأنشطة الفكرية، كحل المشكلات، واتخاذ القرارات على سبيل المثال. والأهم من ذلك، أن الذكاء الاصطناعي من المفترض أن ينطوي على مكون أو عنصر قادر على البرهنة أو إقامة الحجة، كما في مقدور البشر، في المواقف التي لا تتاح فيها سوى المعرفة الجزئية أو الارتباط غير المباشر؛ فنظام الذكاء الاصطناعي لأغراض تنظيم المعلومات، ينبغي أن يكون قادرا على إيجاد مصطلحات أوسع أو أضيق أو متصلة، تبعا لمحتوى الوثائق. وبعبارة أخرى، فإن النظام ينبغي أن يكون قادرا على التحقق من العلاقات بين المصطلحات التي تتضمنها الوثائق وإقرار هذه العلاقات. وينبغي للنظام أن يتمتع بذكاء اصطناعي، يحاكي الذكاء البشري، لكي ينجز المهمة.

ويتم الحصول على المعرفة، في الذكاء الاصطناعي، عن طريق عملية هندسة المعرفة، التي يحصل فيها مهندس المعرفة على بغيته من خبير المجال، وهو إنسان واسع المعرفة إلى أبعد حد، في حقل أو مجال بعينه، وذلك اعتمادا على طرق كالمقابلة، والملاحظة، وأساليب التفكير على نحو مسموع. ثم يتم تمثيل المعرفة المكتسبة أو التعبير عنها اعتمادا على قواعد الإنتاج(*) **production rules** أو الشبكات الدلالية، أو الأطر، على سبيل المثال لتشكيل قاعدة معرفية: وتشكل القاعدة المعرفية، إلى جانب التدابير الأخرى (كمحرك الاستدلال **inference engine**، وإمكانية التعلم الذاتي، على سبيل المثال) أحد تطبيقات الذكاء الاصطناعي.

(*) ناتج برنامج حاسبي، يستخدم عادة لتوفير شكل ما من الذكاء الاصطناعي، يتكون في المقام الأول من مجموعة من القواعد حول السلوك. وهذه القواعد تمثل أساس ثبتت جدواه في التخطيط الآلي، والنظم الخبيرة. (المترجم)

وقد تمت تجربة طرق مختلفة، في بحوث الذكاء الاصطناعي، في العقود الماضية. وقد أشار مولنز **Mullins (2005)** إلى الأساليب التالية، القائمة على المعرفة، بالنسبة للذكاء الاصطناعي، عندما عاود النظر في موضوع الآلات القادرة على التفكير؛ وهي المنطق الرمزي، والشبكات العصبية الاصطناعية، والاستدلال الافتراضي **Bayesian**، والتصوير بالرنين المغناطيسي الوظيفي **functional magnetic resonance imaging (fMRI)**.

والمنطق الرمزي نهج رياضي، يتم فيه التعبير عن الأفكار والمفاهيم، برموز كقواعد الإنتاج أو التسلسلات الهرمية للمفاهيم. ثم تعالج هذه الرموز فيما بعد، وفقا لقواعد المنطق. ونظرا لتوافر المعلومات الكافية؛ هناك أمل في أن تصبح نظم المنطق الرمزي هذه، ذكية في النهاية. وفي شرح أساليب التعلم الرمزي للآلات، وهي مكون لا غنى عنه في المنطق الرمزي، أشار تشن **Chen (1995)** إلى أن التعلم من الأمثلة، وهو حالة خاصة للتعلم الاستقرائي **inductive**، يبدو أكثر أساليب تعلم الآلات رمزيا، محملا بالوعود، بالنسبة لاكتشاف المعرفة، أو تحليل البيانات؛ فالمستفيد على سبيل المثال، يسترجع مجموعة مبدئية من المقالات، ويقسمها إلى صالحة وغير صالحة بالنسبة لاهتمامه. ثم تستخدم هذه الأمثلة الإيجابية والسلبية بعد ذلك لتدريب النظام على التقاط الملامح المشتركة للمقالات الصالحة **(Jenning & Higuchi, 1992)**.

وتحاكي الشبكات العصبية الاصطناعية، في المستوى البدائي، الطريقة التي تعمل بها الخلايا العصبية في المخ؛ فهي تتكون من خلايا عصبية اصطناعية مترابطة، أي البرمجة التي تحاكي خواص الخلايا العصبية البيولوجية. وتحاكي الخوارزميات الوراثية **genetic algorithms**، وهي أحد أشكال تطور الشبكات العصبية، الميراث الجيني والقدرة على استنباط حلول أفضل لمشكلة ما في كل جيل، بينما تكفل الخوارزميات الارتباطية، القدرة على اكتساب المعرفة من جانب إحدى شبكات

الخلايا العصبية المترابطة (نقاط الارتكاز **nodes**) ونقاط الاشتباك أو الارتباط العصبي **synapses** الموزونة (الروابط **links**) والوحدات المنطقية المبدئية. ومن الممكن استخدام خوارزميات التعلم، لضبط أوزان الارتباط، بحيث يكون النظام قادرا على التنبؤ بالأمثلة غير المعروفة، أو تصنيفها على النحو الصحيح **(Chen, 1995)**. واعتمادا على تصنيف المفاهيم القائم على الشبكات العصبية، على سبيل المثال، أعد تشن وآخرون **Chen, et al. (1994)**، قائمة مبدئية بالأفكار والموضوعات المهمة، التي يتم التعبير عنها في تعليقات اللقاءات أو المؤتمرات.

وقد أدت النظم التي تستخدم الأسلوب الرياضي المعروف باسم الاستدلال الافتراضي، إلى ارتفاع مستوى أداء كثير من برامج الذكاء الاصطناعي، إلى حد يكفل إمكانية الاعتماد عليها في العالم الواقعي. ومساعد مشبك ورق ميكروسوفت أوفيس **Microsoft Office paper clip assistant**، أحد التطبيقات المعتمدة على نظم الاستدلال الافتراضي، على الرغم من أنه لم يلق استقبالا حسنا. ومن ناحية أخرى تبنى لاو وكويرا **Lau and Coiera (2006)** الأنموذج الافتراضي، للتنبؤ بتأثير البحث في العنكبوتية العالمية، على اتخاذ مجموعة من الأطباء للقرارات.

والتصوير بالرنين المغناطيسي الوظيفي **(fMRI)** أسلوب يستخدم لمعرفة أي أجزاء المخ ينشط عندما يكون الشخص بصدد التفكير في شيء محدد بعينه؛ فعندما يطلب من الأشخاص تخيل أداة (مطرقة على سبيل المثال) أو مبنى (منزل على سبيل المثال) فإن مناطق المخ نفسها هي التي تنشط على النحو الذي يمكن أن يحدث عند مشاهدة صورة لهذه الأشياء. ومن المفترض أن يكون بإمكان باحثي الذكاء الاصطناعي، تدريب برنامج للنظر في صور المخ، وأن يقرر بنسبة دقة ٩٠ بالمئة، ما إذا كان ذلك الشخص يفكر في أداة أو مبنى **(Mullins, 2005)**. ويمكن لمثل هذا البرنامج أن يقرأ في النهاية مخ الشخص، وبذلك يجعل بالإمكان التحقق من احتياجات البشر من المعلومات، على نحو أفضل وأكثر وضوحا أو صراحة.

ومما لاشك فيه، أن هـذه الطـرق الأربـع للبحـث في الـذكاء الاصـطناعي، لا تشكل جميع الخيارات المتاحة. وما النهج اللغوي الـذي يطبـق عـلى نطـاق واسـع، في معالجة اللغة الطبيعية، في الذكاء الاصطناعي، إلا أحد أمثلـة هـذه الخيـارات. وبيـنما يتجـاوز حـدود مجـال هـذا الكتـاب، تقـديم دراسـة تحليليـة كاملـة لطـرق الـذكاء الاصطناعي، فإنه يبدو أن باحثي الذكاء الاصطناعي، بصرف النظـر عـن الأسـاليب المحددة التي يمكن لهم دراستها، يتطلعون جميعـا لتطـوير تطبيقـات لمجـالات كثـيرة (كالشخوص الآلية **robotics**، والتحقـق مـن الأنـماط عـلى سـبيل المثـال) وتنظـيم المعلومات واسـترجاعها أحـد هـذه المجـالات. وقـد بـذلت جهـود في مجـال تنظـيم المعلومـات واسـترجاعها، لتطبيـق الـذكاء الاصـطناعي في اقتطـاف أو اقتبـاس **extraction** المعلومات، والتلخيص الآلي، والـرد عـلى الاستفـسارات، والبحـث في اللغة الطبيعية، والوكلاء الأذكيـاء **intelligent agents**، فـضلاً عـن أمـور أخـرى. وتبـدو معالجـة اللغـة الطبيعيـة والعنكبوتيـة الدلاليـة، همـا مجـالا البحـث في الـذكاء الاصـطناعي اللـذان يمكـن أن يـسهما عـلى أحـسن وجـه، في مجـال تنظـيم المعلومات واسترجاعها.

معالجة اللغة الطبيعية:

معالجة اللغة الطبيعية مجال للبحث والتطبيق، يستكشف كيف يمكن لنصوص اللغة الطبيعية، التي يتم إدخالها للحاسب، أن تعالج وتحـول إلى شـكل أكثـر ملاءمـة للمزيد من المعالجة (**Chowdhury, 1999**). وكما ذهب ليدي **Liddy (1998)**، فإن معالجة اللغة الطبيعيـة، مجموعـة مـن الأسـاليب الحاسبية الخاصـة بتحليـل وتمثيـل النصوص، التي ترد على نحو طبيعي، في مستوى واحد أو أكثـر للتحـصيل اللغـوي؛ وذلك لأغراض تحقيق معالجة للغة أشبه بالمعالجة البشرية، من أجل إنجاز مجموعة من

المهام أو التطبيقات المحددة. وتشمل هذه المهام أو التطبيقات في الأساس، التلخيص الآلي، والرد على الاستفسارات (Carbonell et al., 2000). ونناقش هنا كلا من التطبيقين على حدة، بعد بيان دور معالجة اللغة الطبيعية في تنظيم المعلومات واسترجاعها:

دور معالجة اللغة الطبيعية في تنظيم المعلومات واسترجاعها:

اللغة المنطوقة أو المكتوبة، مكون أساس في كل من تنظيم المعلومات واسترجاعها؛ نظرا لأن المعلومات ينبغي التعبير عنها، عندما يتم تجهيزها ونقلها أو إيصالها. واللغة الطبيعية، في تنظيم المعلومات واسترجاعها، في مقابل أي لغة اصطناعية أخرى، هي اللغة التي يمكن للبشر استعمالها للتعبير عن المعلومات أو تمثيلها، أو صياغة الاستفسارات. ومن الممكن معالجة اللغة الطبيعية في الذكاء الاصطناعي، في المستويات الصوتية، والصرفية، والمعجمية، والنظمية **syntactic**، والدلالية، والحديث أو الخطاب **discourse**، والوظيفية أو النفعية **(Doszkocs, 1986; Liddy, 1998)**، ولكل هذه المستويات مضامينها بالنسبة لتنظيم المعلومات واسترجاعها.

وينطوي المستوى الصوتي للتحليل على معالجة أصوات الكلام. ويمكن لهذا الضرب من المعالجة أن تكون له أهميته في استرجاع الأصوات (راجع ٩/ ٢/ ٢)، الذي يشتمل، على سبيل المثال، على المدخلات الصوتية أو المعلومات المنطوقة. وخاصية "شبه الصوت **sound like**"، في استرجاع المعلومات (كالعثور على سبيل المثال، على الوثائق التي تشتمل على مصطلحات لها جرس مثل "الموسيقى")، لا يمكن تنفيذها إلا على المستوى الصوتي لمعالجة اللغة الطبيعية **NLP**.

ويتعلق المستوى الصرفي للمعالجة بالعناصر التي تستعمل لتكوين الكلمات المفردة، كجذوع الكلمات أو جذورها، والسوابق، والحشو، واللواحق، والكلمات المركبة؛ فكلمة **uniterm**، على سبيل المثال، تتكون من السابقة **uni**، وجذر الكلمة **term**. والبتر أسلوب للاسترجاع مشتق من هذا المستوى للمعالجة. ومن الممكن لتنظيم المعلومات، كالتكشيف الآلي على سبيل المثال، أن يتم أيضا، وبدرجة معينة، عند معالجة الوثائق على المستوى الصرفي.

وعادة ما تدل المعالجة على المستوى المعجمي **lexical**، على مغزى الكلمات واستعمالها، كتحديد ما إذا كانت كلمة **book** على سبيل المثال، قد قصد بها كما في العمل المنشور أم كما في إجراء حجز. وتشمل المضامين بالنسبة لتنظيم المعلومات واسترجاعها، في هذا المستوى للتحليل اللغوي: (١) إزالة أو إزاحة كلمة الاستبعاد **stop-word**، (٢) استبدال مصطلح البحث الآلي وإضافته (كالبحث في المكنز على سبيل المثال)، (٣) اكتشاف الأخطاء الهجائية وتصحيحها، (٤) معالجة الأسماء الاستهلالية والمختصرات عن طريق عمليات البحث في قوائم المفردات.

ويعرب المستوى النظْمي للتحليل الجمل؛ لتوضيح ما بين كلمات الجمل من علاقات نحوية؛ ما إذا كانت الوحدة البنيوية جملة أم عبارة أم فقرة، على سبيل المثال. وتنطوي هذه العملية أيضا على الحاجة إلى (النحو، الذي يصف الترتيب المحتمل للكلمات ومختلف قواعد التوافق؛ فمن الممكن للعبارات الإسمية على سبيل المثال، التي يتم التحقق منها في التحليل على المستوى النظْمي، أن تكون مرشحة كمصطلحات للتكشيف. والبحث بالعبارات، والبحث بالتجاور، كل هذه تطبيقات للغة التي تعالج على المستوى النظْمي.

وتحاول المعالجة على المستوى الدلالي إضفاء معنى على مختلف المكونات النظمية للجملة. ومن الممكن أن تدعو الحاجة إلى مختلف المساعدات الدلالية

(كالمكانز على سبيل المثال) لإجراء التحليل اللغوي في هذا المستوى. وتتكون عينة للتطبيقيات في هذا الصدد بالنسبة لتنظيم المعلومات واسترجاعها، من العرض الآلي للإحالات، والمترادفات، والمصطلحات المتصلة. إلا أنه لا يمكن للمعالجة على المستوى الدلالي، أن تتم بنجاح دون التحليل اللغوي على المستوى الوظيفي.

ويفسر مستوى الخطاب **discourse** لمعالجة اللغة (الذي لم يذكر في **Doszkocs, 1986**) المبنى والمعنى اللذين تحملهما النصوص الأكبر من الجملة. وبعبارة أخرى، فإنه يمكن لمثل هذه المعالجة أن تكفل القدرة على إيصال المعنى على نحو يتجاوز أو يخرج عن نطاق ما يمكن إيصاله بالكلمات أو الجمل المفردة. ويمكن للغة الطبيعية، التي تعالج على مستوى الخطاب، أن تستعمل لفهم الدور المحدد لكل عنصر من عناصر المعلومات في الوثيقة؛ ما إذا كان، على سبيل المثال، خلاصة، أو وجهة نظر، أو تنبؤ أو حقيقة ببساطة. ومن الواضح أن لهذا البنيان مضامينه المحددة بالنسبة لاسترجاع المقاطع أو الفقرات **passage**؛ نظرا لأن المقطع في هذا السياق، يمكن أن يكون أطول من الجملة.

ويحاول المستوى الوظيفي **pragmatic**، أعلى مستويات معالجة اللغة، تحديد معنى اللغة على ضوء السياق المحيط، والمؤلف، والمستفيد، والإحاطة بالعالم الواقعي. ويمثل هذا المستوى من المعالجة أكبر تحدٍ لباحثي الذكاء الاصطناعي، في معالجة اللغة الطبيعية. فالبشر يعتمدون فعلا على السياق والمعارف الأخرى لفهم اللغة والتخلص من الغموض. إلا أن ما تنطوي عليه هذه العملية من تعقد وصعوبة، لم يتم استيعابه كاملا بعد، ولم تتطور النظم القادرة على محاكاة كيفية معالجة البشر للغة في هذا المستوى. ويمكن لتطبيق هذه العملية في تنظيم المعلومات واسترجاعها، أن يكون أمرا غير عادي، وعلى قدر هائل من الأهمية، إذا ما أمكن تحقيق اختراق في معالجة اللغة الطبيعية، في هذا المستوى. ويمكن لكثير من التطبيقات التي كانت تراود الأحلام، في

تنظيم المعلومات واسترجاعها، كالتلخيص الآلي، والاستفسار باللغة الطبيعية، أن تصبح حينئذ حقيقة واقعة. كذلك يمكن لتبديل اللغات **vocabulary switching**، الذي عرضنا له بإيجاز في نهاية الفصل الرابع، كوسيلة للانتقال تلقائيا، من إحدى لغات تنظيم المعلومات واسترجاعها إلى أخرى، عبر مجالات موضوعية مختلفة، أن يفيد أيضا من اتجاه البحث هذا في الذكاء الاصطناعي.

وإذا ما أمكن إنجاز معالجة اللغة بنجاح على المستوى الدلالي والوظيفي والخطاب، فإنه يمكن للعنكبوتية الدلالية التي تصورها تيم بيرنرز-لي **Tim Berners-Lee** ورفاقه، أن تصبح حقيقة واقعة. ونستكشف موضوع العنكبوتية الدلالية في القسم ١٢/ ٣. ومن ناحية أخرى، تحدد المستويات السبعة المختلفة التي تنطوي عليها معالجة اللغة الطبيعية، معالم ما يمكن لمعالجة اللغة الطبيعية أن تحققه بالنسبة لتنظيم المعلومات واسترجاعها. وكلما ازدادت وحدة التحليل حجما (أي من المورفيم **morpheme**، أي أصغر وحدة صرفية، إلى الكلمة، ثم الجملة، ثم الفقرة، إلى الوثيقة الكاملة) تتراجع دقة الظواهر اللغوية، ويزداد الخيار الحر والقابلية للتغير. وبينما يمكن إنجاز المستويات الأربعة الأولى للمعالجة، على نحو معقول في أيامنا هذه، فإن الحاجة تدعو لإجراء المزيد من البحوث في إتمام التحليل اللغوي على المستوى الدلالي، ومستوى الخطاب، وخصوصا المستوى الوظيفي. ويمكن لجهود البحث في علم اللغة الحاسبي، واللغويات بوجه عام، وفلسفة اللغة، وعلم اللغة الاجتماعي، أن تيسر بحوث الذكاء الاصطناعي، في معالجة اللغة الطبيعية **(Cercone & McCalla, 1984)**.

ونظرا لأهمية اللغة في تنظيم المعلومات واسترجاعها؛ فإن مضامين معالجة اللغة الطبيعية بالنسبة لتنظيم المعلومات واسترجاعها، تبدو واسعة المدى. ويشكل التلخيص الآلي، والرد على الاستفسارات (**QA**) المجالين التطبيقيين الرئيسين في هذه

الفئة (Carbonell, et al., 2000)، بينما يمكن النظر إلى التطبيقات الأخرى (كاقتطاف المعلومات أو اقتباسها، والبحث في اللغة الطبيعية، على سبيل المثال)، بوصفها امتدادا لما ينطوي عليه التطبيقان الرئيسان أو اشتقاقاً منهما. ونناقش في أقسام لاحقة، التلخيص الآلي والرد على الاستفسارات، ونتبعهما بتطبيقات الذكاء الاصطناعي الأخرى في مجال تنظيم المعلومات واسترجاعها.

التلخيص الآلي:

التلخيص الآلي هو إعداد طبعة موجزة من نص ما، بأحد برامج الحاسب. وترجع جهود البحث في هذا الصدد، إلى خمسينيات القرن العشرين، حينما كان هانز بيتر لون **Hans Peter Luhn** يجري تجاربه على إعداد المستخلصات آليا **(1958)**. ولما كانت المعلومات في العصر الرقمي، تنمو بمعدلات أسية، وتصل إلى مستوى بيتا **Peta** (أي ١٠ ١٥) في وقت قصير جدا **(Campbell, 2008)**، استخدمت طرق مختلفة، كالتلخيص الآلي على سبيل المثال، للتعامل مع سيول البيانات الجارفة. والتلخيص الآلي، الذي يبدو في سبيله في الواقع لأن يصبح قطاعا في حد ذاته، في مجال تنظيم المعلومات واسترجاعها، نظام يهدف على وجه التحديد، لإنتاج ملخصات، باستعمال تقنيات معالجة اللغة الطبيعية.

وللتلخيص في هذا السياق معنيان؛ أولهما الاقتطاف أو الاقتباس **extraction**، وثانيهما الاستخلاص **abstraction**، وإن كان ينبغي الإشارة إلى أن هذا الأخير ليس هو نفسه المستخلص التقليدي، الذي عالجناه في ٢/ ٣/ ١/ ١. فكما ذهب ماني **Mani** **(2001)**، فإن الاقتباس ملخص يتكون برمته، من مواد مستنسخة من المصدر الذي يتم إدخاله، بينما المستخلص ملخص يشتمل على الأقل، على بعض العناصر التي لم يكن لها وجود في الأصل الذي تم إدخاله. وتشمل المحددات الأخرى للتلخيص الآلي معدل الضغط أو التركيز **compression** (أي نسبة طول الملخص إلى طول

المصدر الأصلي)، والمتلقي (أي ما إذا كان عاما أم مركزا في اهتماماته)، وحجم المصدر (أي ما إذا كان وثيقة واحدة أم عدة وثائق)، والجنس الأدبي **genre** (كالأخبار، أو رسائل البريد الإلكتروني، أو رسائل المدونات، أو صفحات العنكبوتية)، واللغة (أي ما إذا كان أحادي اللغة، أم متعدد اللغات، أم عبر اللغات) والوسائط (كالنصوص، والأصوات، والصور، والفيديو). وعادة ما تصنف الملخصات وفقا لهذه المحددات؛ فالملخصات التي تنشأ للجمهور العام، على سبيل المثال، تسمى الملخصات العامة.

عملية التلخيص الآلي:

كما يمكن أن يتبين لنا من قائمة المحددات التي سبق ذكرها، فإن بعض أنواع الملخصات الآلية (كتلك الخاصة بالوثائق المفردة على سبيل المثال) أيسر في الحصول عليها، من الأنواع الأخرى (كتلك الخاصة بالوثائق المتعددة على سبيل المثال). وتنطوي عملية التلخيص الآلي، بوجه عام، على الخطوات الرئيسة التالية: التحقق من الجملة واختيارها، واقتطاف الجملة، وعمليات صقل **smoothing** الملخص. وفي الوقت الذي لا يمكن فيه الاستغناء عن الخطوتين الأوليين بالنسبة لأي تلخيص آلي، فإن عمليات صقل الملخص قد لا يتم إنجازها؛ نظرا لأن معظم حالات التلخيص الآلي تتم بمجرد ترتيب الجمل، بدلا من صقل الملخص في هذه الخطوة. ولا تتم عمليات الصقل الحقيقية للملخص، أي جعل الملخص أكثر قابلية للقراءة، لا تتم إلا عندما يكون الناتج النهائي مستخلصا **(Mani, 2001)**. وبإمكان النظم الراهنة للتلخيص الآلي، بوجه عام، إعداد مقتطفات أو اقتباسات لا مستخلصات **(Over, Dang & Harman, 2007; Spärck Jones, 2007)**. ومن بين الخطوات الثلاث التي تم التحقق منها في عملية التلخيص الآلي، يعد التحقق من الجملة واختيارها الأهم والأصعب، في حين يتم اقتطاف الجملة آليا في المقام الأول، ولا يتم صقل الملخص إلا بالنسبة للمستخلصات فقط، ومن ثم فإن أنموذج تجهيز الملخص

التخطيطي **schematic** للنص، الذي عرضته سبارك جونز **Spärck Jones (2007)**، في الشكل رقم (١) في تقريرها، يبدو أكثر ملاءمة بالنسبة لإنتاج المستخلصات. والأمر كذلك أيضا بالنسبة للهندسة المعمارية **architecture** للملخص رفيع المستوى، الذي صوره ماني **Mani (2001)**.

المقاربات الرئيسة للتلخيص الآلي:

كما هو الحال بالنسبة لأنشطة تنظيم المعلومات واسترجاعها الأخرى، هناك مقاربات أو مناهج مختلفة للتلخيص الآلي. وقد قسمت سبارك جونز **Spärck Jones (2007)** هذه المقاربات إلى اقتباسية **extractive** في مقابل غير الاقتباسية، وذلك بناء على ما إذا كان الناتج النهائي اقتباسا. ثم تحققت بعد ذلك من ثلاث فئات فرعية للمقاربات الاقتباسية؛ هي الإحصائية الأساسية، والإحصائية المدعمة **enriched**، وتعلم الآلات **machine learning**. وتدل الطرق الإحصائية الأساسية للتلخيص الآلي، على استعمال تردد المصطلحات **tf**، والتردد العكسي للوثائق **idf** و(راجع ٥/ ١/ ٢/ ٢)، أو خوارزميات الوزن الأخرى، للترتيب الطبقي للجمل المصدرية. ويتم اختيار الجمل من القائمة الطبقية لأعلى الأرصدة **top scores**، حتى الوصول إلى الحد الفاصل **threshold** لطول الملخص (كأن يكون ١٠ بالمئة من طول المصدر على سبيل المثال). وعادة ما يتم إيصال الجمل التي يقع عليها الاختيار، بالترتيب نفسه، الذي وردت به في المصدر الأصلي. وتشمل أوجه قصور مثل هذه الملخصات، ترتيب الجمل كما هي في المصدر، وتحديد وضعها الطبقي بناء على خوارزميات قائمة على التردد فقط.

والمقاربات الإحصائية المدعمة، مصممة إلى حد ما، للتغلب على أوجه قصور الطرق الإحصائية الأساسية. ويتم تدعيم هذه الأنواع من الاستراتيجيات، عن طريق الوحدات والخواص المعجمية، أو من خلال بنيان المصدر. وفي الطريقة الأولى

للتدعيم، يتم إجراء معالجة معقدة للعناصر المعجمية، التي يتم حساب إحصاءاتها. وتشمل هذه المعالجة كلا من أنواع الوحدات التي يقع عليها الاختيار (كالكلمات المتصاحبة **co-words**، أو مجموعات الكلمات) والوزن التفاضلي لأنواع الوحدات. كذلك يتم أيضا التحقق من العناصر المعجمية المحددة ذات الخصائص التي تدل على الأهمية (كالخلاصات على سبيل المثال)، والكيانات الفريدة لمصادر العنكبوتية (كالروابط والمحددات الموحدة لأماكن المصادر **URLs** على سبيل المثال) في عملية التلخيص. وفي مقابل ذلك تعتمد الطريقة الأخرى للتدعيم، على الاستخدام المكثف لبنيان المصدر، وذلك بالخروج عن حدود الجمل، ومراعاة معلومات من قبيل بنية الخطاب على سبيل المثال.

ويمثل تعلم الآلات الفئة الفرعية الثالثة للمقاربات الاقتباسية للتلخيص الآلي، فمن المفترض في التلخيص، إذا ما تم على النحو الصحيح، أن يكون تعبيرا عما يحدد مواصفات المصدر. ومن ثم فإنه من الطبيعي التحقق مما إذا كان التلخيص عن طريق تعلم الآلات، يمكن أن ينتقي الملامح المناسبة للمصدر، بما في ذلك أوزان الملامح، ومجموعات الملامح، من بين مجموعة الخيارات المحتملة. ويمكن لتعلم الآلات، وهو أسلوب غالبا ما يطبق في بحوث الذكاء الاصطناعي - يمكن أن ينقسم إلى التعلم المراقب **supervised**، والتعلم غير المراقب. ويكون التعلم مراقباً إذا ما كان الأمر يتطلب تدخلا ما (كتوفير أمثلة التدريب) من جانب المستفيد، من أجل عملية التعلم (**Turmo, Ageno & Gatala, 2006**). وكان الباحثون فيما مضى يستخدمون تعلم الآلات، في التلخيص الاقتباسي لتوجيه الاقتباس، والتحقق من علاقات الخطاب، وترتيب الجمل المقتبسة، وغير ذلك (**Spärck Jones, 2007**).

وعلى عكس الطرق الاقتباسية التي سبقت مناقشتها، فإن الطرق غير الاقتباسية قد تطورت لإعداد مستخلصات في التلخيص الآلي. وينبغي أن تكون مثل

هذه المستخلصات، التي يفترض أن تمثل النص المصدري، بما لا يتجاوز حدود العناصر المعجمية المفردة، تعبيراً أفضل عن المصدر. إلا أنه لم يكن هناك سوى قدر ضئيل نسبيا من التلخيص غير الاقتباسي في العقد الماضي (**Spärck Jones, 2007**)، الأمر الذي يدل بدوره، إلى حد ما، على ندرة الأساليب المجدية الناجحة، لإنتاج ملخصات غير اقتباسية. وفضلا عن ذلك، عادة ما ينطوي ما يتم إنتاجه في هذه الفئة من الملخصات، على درجة ما من إعادة صياغة للمحتوى المصدري (**Mani, 2001**) وليس تمثيلا موجزا لهذا المحتوى.

ومن منظور المعالجة اللغوية، قسم ماني (**Mani (2001**) مقاربات التلخيص على نحو يختلف عما ذهبت إليه سبارك جونز (**Spärck Jones (2007**)، على الرغم من أن هناك أيضا ثلاثة أنواع من الطرق أو المقاربات؛ وهي المقاربات السطحية **shallow**، والمقاربات المتعمقة، والمقاربات الهجين. ولا تحاول المقاربات الضحلة أو السطحية للتلخيص، الخروج عن حدود المستوى النظْمي لمعالجة اللغة الطبيعية، بينما تتكفل المقاربات المتعمقة، بوجه عام بالمستوى الدلالي للجملة على الأقل، والمقاربات المتعمقة للتلخيص الآلي أكثر صعوبة من المقاربات السطحية في التطبيق، نظرا للوضع الراهن للمعرفة في تقنيات معالجة اللغة الطبيعية. وتسمى المقاربة الثالثة للتلخيص، التي حددها ماني (**Mani (2001**، بالطريقة الهجين، وتنطوي على قدر كبير من إعادة ترتيب النص، على نحو يشمل تغيير العلاقات المعتمدة على الخطاب، كجزء من عمليات صقل الملخص. وكما سبق أن بينا فإن عمليات صقل الملخص، لا تتم بانتظام أثناء التلخيص، ويرجع ذلك إلى حد ما، إلى أن التقنيات الخاصة بمثل هذه العمليات لا تزال في مرحلة التطوير.

تقييم التلخيص الآلي:

لما كان التقييم أمراً لا غنى عنه للارتفاع بمستوى جهود تنظيم المعلومات واسترجاعها، فإن مؤتمر فهم الوثائق **Document Understanding Conference (DUC;duc.nist.gov)**، وهو برنامج، يشبه سلسلة مؤتمرات استرجاع النصوص **TREC series**، من أوجه كثيرة، فيما عدا تركيز مؤتمر فهم الوثائق، الذي يعقد منذ عام ٢٠٠١، على تلخيص النصوص. وقد نظر أوفر ودانج وهارمان **Over, Dang & Harman (2007)**، بعمق في عمليات التقييم، التي تمت في مؤتمر فهم الوثائق، من عام ٢٠٠١ حتى عام ٢٠٠٦ **DUC 2001-2006**، إلى جانب الأنشطة المماثلة، التي تتم في مؤتمر تقييم التلخيص **Summarization Evaluation Conference (SUMMAC)**، وإن تي سي آي آر **NTCIR** (إن آي آي مجموعة الاختبار لنظم استرجاع المعلومات **NII Test Collection for IR Systems**). وقد استخدم نوعان رئيسان من التقييم في هذين المؤتمرين، من أجل اختبار الملخصات؛ وهما التقييم الجوهري أو المتعمق **intrinsic**، والتقييم السطحي **extrinsic**. ويركز التقييم الجوهري على وجه التحديد، على قياس جودة الملخصات الناشئة مباشرة، بينما يركز التقييم السطحي على قياس مدى نجاح الملخصات في دعم الأداء في مهام معينة. وموضوع مؤتمر فهم الوثائق **DUC** هو تلخيص النصوص الذي يشمل التلخيص الآلي، ولكنه يشمل أيضا الطرق اليدوية في العمل. وبالإضافة إلى الكشف عن أن الملخصات الآلية الناشئة في مؤتمر فهم الوثائق من عام ٢٠٠١ إلى ٢٠٠٦، مقتطفات في المقام الأول، سجل أوفر ودانج وهارمان **Over, Dang & Harman (2007)**، النتائج التالية:

- معظم الملخصات اليدوية أفضل ولاشك من معظم الملخصات الآلية.
- لا تختلف معظم الملخصات الآلية اختلافا جوهريا عن بعضها البعض.

- عادة ما تختلف الملخصات الآلية في حدودها القصوى، اختلافا جوهريا فيما بينها.

- .نادرا ما يكون أداء الملخصات الآلية أفضل من الحدود الدنيا (أي التقاط الجمل أو الفقرات الافتتاحية) اعتمادا على بنية المقالات الإخبارية.

وفضلا عن ذلك، اقترح أوفر ودانج وهارمان **Over, Dang & Harman (2007)** ضرورة انتقالِ مؤتمرات فهم الوثائق **DUC** المستقبلية، لا إلى ما يتجاوز المواد الإخبارية فحسب، وإنما إلى مهام التلخيص الجديدة، التي توحي بها أوجه الإفادة الحقيقية من الملخصات (كالملخصات في مجالات بعينها كالطب والقانون على سبيل المثال).

وقد تطور مؤتمر تحليل النصوص **Text Analysis Conference (TAC,www.nist. gov/tac)**، الذي بدأ عام ٢٠٠٨، عن مؤتمر فهم الوثائق، ومسار الرد على الاستفسارات **(QA)** بمؤتمرات استرجاع النصوص **TREC**. ويهدف مؤتمر تحليل النصوص إلى مساندة البحث، في نطاق أوساط معالجة اللغة الطبيعية، وذلك بتوفير البنية الأساس اللازمة للتقييم واسع المدى، للمناهج الخاصة بمعالجة اللغة الطبيعية. وبرعاية المعهد الوطني للمواصفات المعيارية والتقنية **NIST**، وغيره من الأجهزة الحكومية الأمريكية، يحل مؤتمر تحليل النصوص محل مؤتمر فهم الوثائق في الأساس، بينما ينهض بأدوار أخرى (كتقييم الرد على الاستفسارات على سبيل المثال) في مجال معالجة اللغة الطبيعية.

ويمكن للتلخيص الآلي أن يكون أي شيء إلا أن يكون موضوعا بسيطا، نظرا لتعدد أبعاده **(Carbonell et al., 2000)**. وتسهم المتغيرات التي تناولها ماني **Mani (2001)** (كمعدل الضغط أو التركيز، والمتلقي، وحجم المصدر، والجنس الأدبي،

ونوع الملخص) جميعها في هذا التعدد في الأبعاد، الذي يدل في النهاية على ما يكتنف التلخيص الآلي من صعوبة وتعقد. ومن ناحية أخرى فإن البحث في التلخيص يتسع مداه، كما يحرز التقدم (**McDonald & Chen, 2006** على سبيل المثال). ويمكن للإنجازات المستقبلية في التلخيص الآلي، أن تيسر إلى حد بعيد، الكثير من أنشطة تنظيم المعلومات واسترجاعها.

الرد على الاستفسارات:

يمكن لكل من يجري عملية بحث، اعتمادا على نظام للاسترجاع (كأن يكون نظاما على الخط المباشر، أو فهرسا متاحا على الخط المباشر، أو البحث بأحد محركات البحث في العنكبوتية) أن يتلقى بعض النتائج في المقابل. ويمكن لمثل هذه النتائج أن تشتمل على الإجابة التي يلتمسها المرء أو لا تشتمل، وعلى المستفيد أن يتصفح النتائج. وبعبارة أخرى، فإن نظام استرجاع المعلومات، لا يمكنه سوى الحصول على بعض النتائج، وعرضها للمستفيد، بينما يتعين على المستفيد أن يمعن النظر في النتائج، للعثور على الإجابة إذا كانت هناك فعلا. ومهمة المستفيد في العثور على الإجابة من النتائج المسترجعة، ليست هينة، لأسباب متعددة (كالوقت، والحكم على الصلاحية، والمصادر التي لا يمكن الوصول إليها، والمجموعات). ومن ثم فقد اقترحت تقنية الرد على الاستفسارات، لإعفاء المستفيد من عبء غربلة الإجابة من نتائج عملية البحث، التي غالبا ما تعد بالملايين في العصر الرقمي.

ويعني الرد على الاستفسارات، بإيجاز أن نظام استرجاع المعلومات قادر على تقديم الإجابة للمستفيد على وجه التحديد، وذلك في نتائج عملية البحث، بدون أي معلومات غير صالحة، تختلط بهذه الإجابة. وقد قدم مايبري (**Maybury (2004**) تعريفا أكثر شمولا للرد على الاستفسارات **QA**:

الرد على الاستفسارات (QA) عملية تفاعلية بين الإنسان والحاسب، تشمل إدراك حاجة المستفيد إلى المعلومات، التي عادة ما يتم الإعراب عنها باستفسار باللغة الطبيعية، ثم استرجاع الوثائق أو البيانات، أو المعرفة، من مصادر منتقاة، واقتطاف وتهيئة وترتيب أولويات الإجابات المتوافرة من هذه المصادر، ثم عرض الاستجابات، وشرحها بطريقة فعالة. (P. 3)

عملية الرد على الاستفسارات:

في الوقت الذي يبدو فيه مفهوم الرد على الاستفسارات جاذبا إلى أبعد حد، فإن تحقيق هذا الهدف ينطوي على تحديات في جميع خطوات العملية. وعادة ما يتناول الباحثون في المجال، كيف يتم الرد على الاستفسارات من منظور معارفهم وخبراتهم (راجع على سبيل المثال **Kwok, Etzioni & Weld, 2001; Lin & Katz, 2006**). وعادة ما يشتمل أسلوب الرد على الاستفسارات، بوجه عام، على إعراب الاستفسارات أو تحليلها، وتصنيف السؤال، وصياغة الاستفسار، واقتطاف الإجابة، واختيار الإجابة، وعرضها.

والخطوة الأولى في عملية الرد على الاستفسارات هي إعراب الاستفسار، أو تحليل الاستفسار لفهم السؤال الذي يثيره. وإذا ما كان نظام استرجاع المعلومات يدعم البحث في اللغة الطبيعية (راجع ١٢/ ٢/ ٤)، فإن هذه الخطوة عادة ما تعالج بواسطة تقنية مدمجة **built-in**. وما لم يكن الأمر كذلك، فإن الاستفسار الذي يتم إدخاله في صندوق البحث الخاص بنظام استرجاع المعلومات، ينبغي أن يتم تحليله قبل أن يكون بالإمكان انتقاله إلى الخطوة التالية في الرد على الاستفسارات.

ويمكن للخطوة التالية أن تكون تصنيف السؤال وفقا لإطار ما. وقد بنيت أطر مختلفة لتصنيف الأسئلة على هذا النحو، على مر السنين. وقد صنف كاربونل ورفاقه **Carbonell, et al. (2000)** كل هذه القضايا في أربعة مستويات (أي السائل العارض **casual**، والسائل المقولب أو النمطي **template**، والمرسل المبتدئ **cub reporter**، ومحلل المعلومات المحترف)، وذلك في الوقت الذي درس فيه بومرانتس **Pomerantz (2005)** خمسة تقسيمات رئيسة للأسئلة (أي أطر لتصنيف الأسئلة) في الرد على الاستفسارات:

١. أحرف دبليو **W** الستة **Six Ws** (أي من **who**، وماذا **what**، ومتى **when**، وأين **where**، وكيف **how**): وينتمي كثير من أسئلة هذا الإطار للفئة الحقائقية **factual**، أو **factoid**، ويبدو تصنيفها على العموم مباشرا. وتشمل أمثلة أسئلة هذا النوع، متى ولد جورج واشنطن؟ وما أعلى مبنى في العالم؟

٢. موضوعات الأسئلة: تصنف الأسئلة في هذا الإطار باستعمال لغات مقننة معينة (مثل تصنيف مكتبة الكونجرس، ومكنز الفنون والعمارة). ويمكن تصنيف هذه الأسئلة وفقا لمحتواها الموضوعي. ويمكن لعينة من هذه الأسئلة أن تكون: ما أكثر طرق التدريس فعالية في التعليم الجامعي؟

٣. أوجه الإفادة من الإجابات المتوقعة: وتصنف الأسئلة في هذا الإطار، بالنظر في وظائف الإجابات المتوقعة أو أوجه الإفادة منها. فهل يمكن، على سبيل المثال، لإجابة للسؤال أن تستعمل لأغراض المقارنة أو التفسير؟ وقد وضع بومرانتس **Pomerantz (2005, Table 1)** تقسيما لأوجه الإفادة من الإجابات المتوقعة.

٤. أشكال الإجابات المتوقعة: ويبدو هذا النهج في التصنيف مماثلا للطريقة التي عادة ما تصنف بها الأسئلة المرجعية التقليدية (كالإرشادية **directional**، والمرجعية الجاهزة على سبيل المثال). كذلك جمع بومرانتس **Pomerantz (2005, Table 2)** الأشكال أو الصيغ المألوفة للإجابات المتوقعة.

٥. أنواع المصادر: تصنف الأسئلة في هذا الإطار وفقا لأنواع **types** (أو أجناس) المصادر التي يمكن أن يتم منها استقاء الإجابات. إلا أن هذا التصنيف لم يطبق لتصنيف الأسئلة في عملية الرد على الاستفسارات **(Pomerantz, 2005)**. وتشمل عينة المصادر كلا من المستخلصات، والكشافات، والمعاجم، والموسوعات، والكتب السنوية أو الحوليات.

وعلى غرار النوع الثالث للتصنيف، لخص بومرانتس **Pomerantz (2005)**، خطة تصنف فيها الأسئلة، بناء على الطريقة التي يمكن بها للسائل (أي المستفيد) أن يفيد من الإجابة التي يسترجعها نظام الرد على الاستفسارات. وقدم كاربونل ورفاقه **Carbonell, et al. (2000)** هذه الخطة. وعادة ما يوجه السائل العارض أسئلة بسيطة وحقائقية. إلا أن السائل النمطي أو المقولب عادة ما يطرح أسئلة يمكن معالجتها بقوالب معيارية، يمكن فيها العثور على أنواع معينة من المعلومات. وتشمل أمثلة هذه الأسئلة: ماذا نعرف عن شركة **ABC**؟، وما قصة حياة الممثل س؟ ويمكن لإجابات الأسئلة النمطية هذه أن تكون أصعب بالنسبة لنظام الرد على الاستفسارات في العثور عليها، مما هو عليه الحال بالنسبة للسائل العارض؛ نظرا لأنه غالبا ما يتطلب الأمر استرجاعها من مصادر متعددة. وتسمى الفئة الثالثة ممن يوجهون الأسئلة بالمراسل المبتدئ **cub reporter**. وقد اختير الاسم ببساطة بوصفه ضربا من التناظر **(Carbonell, et al., 2000)**؛ فالمراسل المبتدئ يركز على العثور على البيانات

الحقائقية، إلا أنه بحاجة أيضا لأن يجمع معلومات من مصادر متنوعة. وفضلا عن ذلك، فإنه ليس هناك قالب معياري متاح في هذه الحالة؛ نظرا لأن بعض المعلومات التي يبحث عنها المستفيد، يمكن أن تكون جوهرية بالنسبة لحاجته، بينما تكون البقية هامشية في أحسن الأحوال. وعلى نظام الرد على الاستفسارات أن يحدد ما ينبغي استرجاعه، اعتمادا على الأساليب المناسبة. وتسمى الفئة الأخيرة في هذه الخطة لتصنيف مقدمي الأسئلة بمحلل المعلومات المحترف، وتمثل أنواعا مختلفة من المستفيدين، على الطرف الأعلى لطيف الرد على الاستفسارات؛ فهؤلاء المستفيدون، كفئة بحاجة إلى إجابات لأسئلة متطورة، على الرغم من أنهم يعملون في مهن مختلفة (ككبار مراسلي الأخبار، والعلماء العاملين على حل المشكلات المتقدمة، والضباط الذين يحللون الاستخبارات على الصعيد الوطني). وغالبا ما يكون الأمر مهمة شاقة، بالنسبة لتطوير نظم الرد على الاستفسارات لهذه الفئة من المستفيدين.

وأيا كان الإطار المستخدم لتصنيف الأسئلة، فإن هذا الإجراء ينبغي اتخاذه قبل الخطوة التالية في الرد على الاستفسارات، وهي صياغة الاستفسار؛ وذلك لضمان تنفيذ الخطوات التالية بناء على ذلك؛ فمن الممكن، على سبيل المثال، للعمليات الخاصة بالحصول على إجابة لسؤال حقائقي، أن تختلف عن تلك الخاصة بالإجابة عن سؤال حول موضوع معين. وفي هذه الخطوة في صياغة الاستفسار، فإن السؤال أو الاستفسار الذي يقدم من جانب المستفيد، يتم تحويله فعلا باستعمال قواعد النظم **syntax** والتفصيلات الأخرى التي يدعمها نظام الرد على الاستفسارات، حتى يمكن معالجته بواسطة محرك بحث النظام على النحو المناسب الصحيح. كذلك يمكن تطبيق إجراءات كتحويل الفعل، وتلقيم الصلاحية المرتد، وصياغة العبارة الاسمية، وذلك عند إعادة صياغة الاستفسار لأغراض الرد على الاستفسارات (راجع على سبيل المثال **Kwok, Etzioni & Weld, 2001**).

والخطوة الرابعة في الرد على الاستفسارات هي استخلاص الإجابة، التي تستقطب بالفعل معلومات معينة من المصادر، وفقا لبعض القواعد. ووراء هذه الخطوة اقتطاف المعلومات (1E). ويمكن للمصادر المستخدمة في عملية اقتطاف المعلومات أن تكون مجموعات يتم بناؤها خصيصا (راجع على سبيل المثال **Lin & Katz, 2006**) أو مجرد ما هو متوافر على العنكبوتية (راجع على سبيل المثال **Neumann & Xu, 2004**). وتعرف القواعد التي يتم الاسترشاد بها في الاقتطاف بالأغلفة **wrappers** التي تصمم عادة لمصدر معلومات بعينه (**Gregg & Walczak, 2006**). ويتم إنشاء الأغلفة، عن طريق الجمع بين تقنيات معالجة اللغة الطبيعية، وتعلم الآلات، والنمذجة الإحصائية، والإلمام بالمجال (راجع على سبيل المثال **Flesca, et al., 2004**). وبمساعدة مثل هذه الأغلفة، يتم التحقق من المحتويات الرئيسة في المصدر (كأسماء الأعلام، والتواريخ، والأماكن) من أجل الاقتطاف (**El Hadi, 2003**). إلا أن وجه الصعوبة الرئيس، في اقتطاف المعلومات **1E** يكمن في توقفه أو اعتماده الجوهري على المجال الموضوعي؛ نظرا لأن المعرفة اللغوية تختلف من مجال لآخر اختلافا بيِّنا. وقد تمت تجربة الضبط اليدوي **manual tuning** للمعرفة اللغوية التي تتوقف على المجال الموضوعي، كمعاجم المصطلحات، والدلالة المعجمية المتصلة بالمجال الموضوعي على وجه التحديد، وأنماط الاقتطاف (**Turmo, Ageno & Gatala, 2006**)، غير أن إجراء اقتطاف المعلومات عبر المجالات الموضوعية، لا يزال يشكل تحديا جوهريا.

ومن الممكن إجراء اقتطاف المعلومات على مستويات متعددة من التجانس **granularity** (**Godly & Reighart, 2001**). فقد حاول كل من موجسيلوفك، وجوميز، وروجوفيتس **Mojsillovic, Gomes and Rogowitz (2004)** اقتطاف الإشارات الدلالية، كالجلد **skin**، والسماء **sky**، والحشائش **grass**، والجليد **snow**،

والأشياء الاصطناعية من مجموعات الصور. وفضلا عن ذلك، فإنه يمكن للاقتطاف لأغراض الرد على الاستفسارات، أن يتم أيضا بمرونة من حيث الطول؛ إذ يمكن لبعض المحتويات التي يتم اقتطافها، أن تكون في حدود كلمة واحدة أو رقم، بينما يمكن لأخرى أن تكون عدة مئات من الكلمات طولا.

واقتطاف المعلومات مكون مهم في استرجاع الفقرات، الأمر الذي يعني أنه ينبغي لاسترجاع الفقرات أن ينطوي على وسيلة للاقتطاف. ومن الممكن النظر إلى استرجاع الفقرات، على نحو ما، بوصفه حلقة في مسيرة تطور نظم الرد على الاستفسارات. ووجه الارتباط بين الرد على الاستفسارات واسترجاع الفقرات، هو أن الأول يقدم للمستفيد إجابات محددة كنتائج، بينما يوحي الآخر فقط بأن المعلومات متضمنة في الفقرة المسترجعة (**Coter, 2005**). وطول النتائج في استرجاع الفقرات، أقل على نحو لا يستهان به، بالمقارنة –بالطبع– بنتائج استرجاع الوثائق.

والخطوة الأخيرة في الرد على الاستفسارات هي اختيار الإجابة بعد اقتطاف المحتوى الأساس الوارد في المصدر. وتماثل هذه الخطوة الحكم على الصلاحية في العملية العادية لاسترجاع المعلومات، اللهم إلا أن البشر، لا النظام، هم من يصدرون الأحكام في أنشطة البحث في غياب تقنية الرد على الاستفسارات. ومما لاشك فيه أنه من الصعب إلى حد بعيد، أن يكون هناك نظام للرد على الاستفسارات، بإمكانه الحكم على الصلاحية؟ نظراً لأن الصلاحية كما بينا في ١١/ ١/ ٢/ ١ مفهوم معقد، إلى حد أن البشر أنفسهم يقفون عاجزين عن التعامل معه على نحو مرضٍ. إلا أن نظام الرد على الاستفسارات لا خيار أمامه له سوى اختيار الإجابات من المحتويات المقتطفة في حد ذاتها. وقد اقترح لن وكاتس **Lin and Katz (2006)** ما يلي، لاختيار الإجابة: فالمحتويات المقتطفة ينبغي أولا أن يتم الحكم عليها؛ ما إذا كانت داعمة أم غير داعمة أي غير صالحة، قبل اتخاذ قرار الاختيار؛ فالمحتويات الداعمة لا تشتمل على خيط

الإجابة فحسب، وإنما ينبغي أن يتم التحقق من الإجابة بثقة، عندما يتم ذلك بواسطة الإنسان، بوصفها صحيحة من المصدر. أما المحتويات غير الداعمة فتشتمل على خيط الإجابة الذي يرد في السياق الصحيح، إلا أنها لا تجيب عن السؤال فعلا. أما المحتوى غير الصالح فإنه إما أن يغفل خيط الإجابة كلية وإما يذكره عرضا أو من قبيل المصادفة. وتحدث هذه الحالة الأخيرة عند البحث بالكلمات المفتاحية، في معظم، إن لم يكن في جميع الأحيان. ومن بين الفئات الثلاث لنتائج الحكم، يمكن اختيار المحتويات الداعمة فقط بوصفها إجابات، وتقديمها للمستفيد، على الرغم من أنه ينبغي أن نلاحظ أن المحكمين البشر هم من يقومون بتحديد المحتويات الداعمة. إلا أن صلاحية الإجابة أمر آخر، يوضع في الحسبان في اختيار الإجابة **(Link & Katz, 2006)**. فإلى أي حد على سبيل المثال، ينبغي أن يكون التاريخ المقصود على هذا النحو؟ فهل العام وحده كاف، أم أنه ينبغي أيضا ذكر الشهر واليوم؟

وبإنجاز خطوة اختيار الإجابة، تصل عملية الرد على الاستفسارات برمتها إلى غايتها، وذلك على الرغم من أن تقديم الإجابة، أحيانا ما يأتي بوصفه خطوة إضافية، أو الخطوة الأخيرة **(Maybury, 2004)**. وكما كان الباحثون يفعلون في التلخيص الآلي، فإنه يتم اختيار مجموعة متنوعة من الطرق (الإحصائية، واللغوية، والقائمة على المعرفة، والتفاعلية على سبيل المثال) في تطوير نظم الرد على الاستفسارات. وتشكل هذه الطرق سلسلة طويلة، وغالبا ما تستخدم بتجميعات مختلفة. وقد وفر مسار الرد على الاستفسارات، الذي استمر من مؤتمر استرجاع النصوص الثامن **TREC-8** حتى مؤتمر ٢٠٠٧، وهو على هذا النحو، واحد من أطول المهام الخاصة وفر منصة قيمة لباحثي الرد على الاستفسارات، للتواصل فيما بينهم. يضاف إلى ذلك أن هذا المسار قد عزز حدوث طفرة في بحوث الرد على الاستفسارات، أسفرت بدورها عن تقنيات جديدة للرد على الاستفسارات **(Voorhess, 2005)**. ومن الممكن ولاشك،

لتأسيس مؤتمر تحليل النصوص TAC، الذي أشرنا إليه بإيجاز في ١٢/ ٢/ ٢/ ٣، أن يوفر المزيد من الدوافع للبحث في الرد على الاستفسارات.

تطبيقات الرد على الاستفسارات:

هناك الآن نوعان من تطبيقات الرد على الاستفسارات؛ ينطوي أحدهما على التدخل البشري في عملية الرد على الاستفسارات، ولا ينطوي الآخر على مثل هذا التدخل. وتعتمد عمليات استرجاع المعلومات في الحالة الأولى، في الأساس، على قاعدة بيانات للأسئلة، إلى جانب قاعدة بيانات للإجابات في المقابل. ويقوم نظام استرجاع المعلومات أولا ببناء قاعدة بيانات الأسئلة، التي يتم تجميعها من مصادر مختلفة (كالاستفسارات التي تشتمل عليها سجلات عمليات البحث **search logs** على سبيل المثال). ثم يتم وضع إجابات لكل سؤال في قاعدة البيانات، اعتمادا على المصادر المتاحة للمستفيدين المستهدفين (كمواقع العنكبوتية على سبيل المثال)، ويتم اختزانها في قاعدة بيانات الإجابات. وعندما يتم إدخال طلب في صندوق البحث، يراجع نظام استرجاع المعلومات قاعدة بيانات الأسئلة الخاصة به، لمعرفة ما إذا كانت هناك مضاهاة لهذا الطلب أم لا. وإذا ما تم العثور على مضاهاة بين طلب البحث وقاعدة بيانات الأسئلة، يسترجع النظام إجابة من قاعدة بيانات الإجابات، باتباع الرابطة الواصلة بين قاعدتي البيانات. وما لم يكن الأمر كذلك، فإنه لا يتم تقديم نتيجة للبحث. وقد استخدمت الطبعة المبدئية لنظام آسك دوت كوم **Ask.com** (الذي كان يعرف من قبل باسم آسك جيفز **Ask Jeeves**) وهو نظام لاسترجاع معلومات الإنترنت، هذه الطريقة الخاصة بقاعدتي بيانات الأسئلة والإجابات، الأمر الذي يوضح لماذا يقدم هذا النظام إجابات كاملة لبعض الأسئلة، ولا إجابات فيما عدا التشويش في معظم الأحيان.

أما الفئة الأخرى من نظم الرد على الاستفسارات (أي تلك التي تعتمد على تقنيات الذكاء الاصطناعي كاملة، وباتخاذ الخطوات التي سبقت مناقشتها) فلا وجود لها؛ نظرا لأن مثل هذه النظم لا تزال دون الوضع الراهن للمعرفة في المجال (**Voorhess, 2005**)، ومن ثم، فإن الباحثين في هذا المجال، يتبعون نهجا للتفرقة والهزيمة **divide-and-conquer** مع الرد على الاستفسارات، بالتعامل مع المهام التي يمكن التعامل معها بشكل أيسر من غيرها (كالرد على الاستفسارات الحقائقية، واقتطاف المعلومات من مصادر العنكبوتية) قبل غيرها. ومن ناحية أخرى، تم وضع خريطة طريق، للتخطيط على نحو نسقي وعلى وجه التحديد، لبحوث المستقبل في الرد على الاستفسارات (**Maybury, 2002**).

ومن الملائم في النهاية، النظر في العلاقة التكاملية المفيدة على وجه الخصوص، القائمة بين التلخيص الآلي ونظم الرد على الاستفسارات؛ نظرا لأن تطور التطبيقات الأخرى (كالبحث في اللغة الطبيعية على سبيل المثال) يتوقف على كليهما. ومن منظور التلخيص، فإن الرد على الاستفسارات أحد طرق كفالة التركيز للتلخيص الموجه نحو الاستفسارات، ومن منظور الرد على الاستفسارات فإن التلخيص أحد طرق اقتطاف المعلومات الصالحة فقط ونشرها، من كل النصوص، إجابة لسؤال غير حقائقي بعينه (**Carbonell et al., 2000**). وفضلا عن ذلك، فإن التلخيص الآلي والرد على الاستفسارات يواجهان، من تحديات جوهرية مماثلة، في مساراتهما نحو النضج، كتطبيقين رئيسين للذكاء الاصطناعي في تنظيم المعلومات واسترجاعها.

البحث باللغة الطبيعية:

يعني البحث باللغة الطبيعية (أو الاستفسار بها) أن بإمكان المستفيد إدخال عمليات بحث محددة (كالعبارة أو السؤال على سبيل المثال) دون المرور بعملية صياغة الاستفسار أو التعبير عنه. وتدل صياغة الاستفسار أو التعبير عنه، الذي نوقش في

٥/ ٣، على عملية ترجمة طلبات البحث، إلى استفسارات يمكن لنظام استرجاع المعلومات معالجتها. فلماذا السماء زرقاء؟ على سبيل المثال، سؤال بحث باللغة الطبيعية، لا يمكن لنظام استرجاع المعلومات التقليدي معالجته على النحو الملائم. إلا أن نظام استرجاع المعلومات نفسه ينبغي أن يكون قادرا على التعامل مع استفسار بحث من قبيل **sky AND blue**، دون أي مشكلات.

ويوضح هذا المثال وجه الاختلاف الجوهري بين نظم استرجاع المعلومات التي تدعم البحث باللغة الطبيعية، وتلك التي لا تدعم ذلك؛ ففي الحالة الأولى يقوم نظام استرجاع المعلومات بإنجاز مهمة صياغة الاستفسار للمستفيد، بينما من المتوقع في الحالة الأخرى، أن يقوم المستفيد بإنجاز عملية صياغة الاستفسار، حتى يكون بإمكان نظام استرجاع المعلومات معالجته. وبعبارة أخرى، فإنه بالنسبة للسؤال "لماذا السماء زرقاء؟" **Why is the sky blue?** كمثال، يمكن لنظم استرجاع المعلومات، القادرة على معالجة استفسارات اللغة الطبيعية، أن تكون قادرة على الأقل، على تنفيذ ما يلي:

- تدل كلمة لماذا **why** على أن ما يتم البحث عنه يتعلق بالسبب، أو العلة، أو إجابات مماثلة أخرى للسؤال.
- أما الكلمتان **is** و **the** الواردتان في السؤال، فهما مما يسمى بالكلمات الوظيفية، ولا يمكن إيرادهما في استفسار البحث.
- أما الكلمتان **sky** و **blue** فمن الكلمات المهمة، حيث ينبغي الربط بينهما بالمعامل البوليني **AND**، ليشكلا استفسارا لهذا السؤال.
- ويتم تجاهل علامة الاستفهام في معالجة الاستفسار باللغة الطبيعية؛ نظرا لأن علامات الترقيم لا تشكل عادة جزءا من الاستفسار، ما لم يتم النص على عكس ذلك على وجه الخصوص.

وربما تبدو هذه الخطوات سهلة بالنسبة للبشر، إلا أنها تفرض تحديات ضخمة بالنسبة لنظام استرجاع المعلومات. وتلك التحديات هي في الأساس ما يحاول الباحثون في معالجة اللغة الطبيعية التصدي لها. وعلى الرغم من أن البحث باللغة الطبيعية يبدو بالفعل شديد الجاذبية، فإن تطبيقه يعتمد بكثافة على البحث والتطوير، في التلخيص الآلي والرد على الاستفسارات، اللذين نوقشا في ١٢/ ٢/ ٢ و١٢/ ٢/ ٣.

ويتم اتباع النهجين الرئيسين لتطوير نظم الرد على الاستفسارات (راجع ١٢/ ٢/ ٣/ ٢) في إجراء البحث باللغة الطبيعية أيضا. وطريقة قاعدتي بيانات الأسئلة والإجابات، التي تناولناها آنفا يكتنفها القصور، إذ أن تقنيتها الأساس يدوية فعلا. وفضلا عن ذلك فإن الطلب لا يمكن بحثه أو النظر فيه، إلا إذا كان متضمنا فعلا في قاعدة بيانات الأسئلة الخاصة بنظام استرجاع المعلومات. غير أنه لا يمكن لأي قاعدة بيانات، مهما بلغت ضخامتها، أن تشتمل على جميع الأسئلة التي يمكن للمستفيدين توجيهها، ومن ثم فإن نظام استرجاع المعلومات يصبح لا جدوى منه، عندما تخلو قاعدة البيانات من سؤال معين، حتى وإن كان النظام يدعم ما يسمى بالبحث باللغة الطبيعية. ويظل النهج الآخر للبحث باللغة الطبيعية، كما أوضحنا في ١٢/ ٢/ ٣/ ٢، رؤية أكثر منه حقيقة؛ نظراً لأن تقنيات معالجة اللغة الطبيعية، بالنسبة لإنشاء نظم عاملة للرد على الاستفسارات، لا تزال في مرحلة التطوير، ومن ثم فإن قليلا فقط من نظم استرجاع المعلومات، التي تدعم البحث باللغة الطبيعية، قد أصبحت حقيقة واقعة حتى الآن.

وأحيانا ما يسمى البحث باللغة الطبيعية بواجهة اللغة الطبيعية؛ نظرا لأن الواجهة هي ما يتفاعل معها المتعامل مع نظام استرجاع المعلومات، أثناء عملية البحث. وفضلا عن ذلك فإن الخطوتين الأوليين في عملية الرد على الاستفسارات (وهما إعراب الاستفسار أو تحليله وتصنيف السؤال) ترسيان في جوهرهما، الأساس

بالنسبة للبحث باللغة الطبيعيـة؛ ولهـذا، فـإن البحـث باللغـة الطبيعيـة، شـأنه شـأن استرجاع الفقرات، مرتبط بتطور تقنية الرد على الاستفسارات. وأخيرا ينبغي أن نبين أن البحث باللغة الطبيعية، ليس بحثا بدون أي استخدام للغات المقيدة (كالمكانز عـلى سبيل المثال). ويسمى البحث في الحالة الأخيرة ببحث الكلمات المفتاحية.

العنكبوتية الدلالية:

العنكبوتية الدلالية **semantic web** هي أكثر المشروعات التـي نفـذها اتحـاد العنكبوتية العالمية **W3 (World Wide Web Consortium)** وكثير من الباحثين، طموحا حتى الآن؛ فقد كانت جزءا من رؤية تـيم بيرنـرز-لي **Tim Berners-Lee**، منذ أعد الطبعة الأولى من العنكبوتية، في مطلع العقد الأخير من القرن العشرين، على الرغم من أن بيانا ووصفا للعنكبوتية الدلالية المتصورة، لم يصدر إلا بعد حوالي عقـد من الزمن في **Berners-Lee, Hendler and Lassila (2001)**. ويعرِّف المؤلفـون في هذا البحث قوي التأثير، العنكبوتية الدلالية بوصفها امتدادا للعنكبوتيـة الراهنـة، يتم فيه إضفاء معنى واضح على المعلومات، على نحو يمكِّـن البـشر والحاسـبات مـن العمل معا بشكل أفضل. وفي سياق تبرير إدخال مجال موضوعي جديـد، وهـو علـم العنكبوتية، في التعليم العالي، واصل بيرنـرز-لي، وهـول، وهنـدلر **Berners-Lee, Hall and Hendler (2006)** القول على وجه التحديد، بأنـه في عنكبوتيـة الوثـائق القابلة للقراءة بواسطة البشر، بإمكان أساليب معالجة اللغة الطبيعيـة اقتطـاف بعـض الدلالة أو المعنى من صفحات نصوص هذه الوثائق. ويمكن لعنكبوتية من البيانـات العلائقية، والتأكيدات المنطقية، والطرق الإيعازية **heuristic**، أن تلخص معا المعـاني أو الدلالات المقصودة التي تستعمل في التواصل البـشري. وفـضلا عـن ذلـك، فـإن العنكبوتية الدلالية، بمجرد تنفيذها، يمكن أن تكفل لـنظم البيانـات التـي وضـعت

بمعزل بعضها عن بعض، الارتباط معا، دون الحاجة إلى اتفاق عام حول المصطلحات والمفاهيم.

ونظرا لأن العنكبوتية الدلالية لا تزال مجرد رؤية، على الرغم من التوصيفات الخيالية المثيرة، التي ترد في نصوص عدة، فإن المزيد من المعالجة النظرية لا يمكن أن يكفل أي تأكيد أو إثبات أو برهان، وإنما يكفل المزيد من التخمين بشأنها. ومن ناحية أخرى، أجريت البحوث النشطة في مجالات عدة، كالهندسة المعمارية للعنكبوتية، لتيسير تطور العنكبوتية الدلالية.

الهندسة المعمارية للعنكبوتية الدلالية:

حتى قبل انتشار عبارة العنكبوتية الدلالية، وضَّح بيرنرز-لي **Berners-Lee** **(2000)** معالم الهندسة المعمارية لهذه العنكبوتية، في رسم بياني من سبع طبقات (الشكل رقم ١٢/ ١). وقد قدمت منذ ذلك الحين، تفسيرات لهذا الرسم البياني، من جانب عدة مؤلفين، في سياقات مختلفة، ومن زوايا مختلفة (راجع على سبيل المثال **Greenberg, 2003; Legg, 2007**). ولا تبدو هناك ضرورة لتكرار شرح كل من الطبقات السبع، في هذه الهندسة المعمارية. وعوضا عن ذلك، سوف نلقي أولاً نظرة فاحصة دقيقة على هذه الطبقات، من القاعدة إلى القمة؛ فالطبقة الثانية (لغة الترميز القابلة للامتداد **XML**+ فضاء الاسم **NS**+ خطة لغة الترميز القابلة للامتداد **xmlschema**)، والطبقة الثالثة (إطار وصف المصادر **RDF**+ خطة إطار وصف المصادر **rdfschema**)، والطبقة الرابعة (لغة التصنيف **ontology vocabulary**)، والطبقات مرتبة في الرسم البياني، تبعا لأهميتها المباشرة بالنسبة لتنظيم المعلومات واسترجاعها. أما الطبقات الأربع الأخرى؛ أي الأولى (الترميز الموحد **unicode** والمحدد الموحد للمصدر **URI**) والخامسة (المنطق)، والسادسة (البرهان) والسابعة (الثقة)، فلا تحظى بالمناقشة في هذا الكتاب؛ نظرا لأن الترميز الموحد والمعرف الموحد

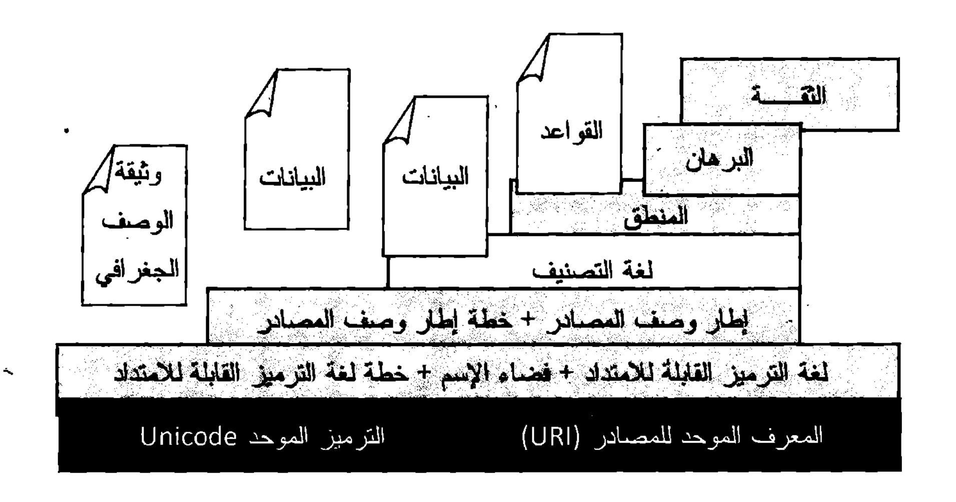
للمصدر يمثلان مرحلة ما قبل العنكبوتية الدلالية، بينما للطبقات الثلاث الأخيرة علاقة غير مباشرة بتنظيم المعلومات واسترجاعها، في هذا السياق، على الرغم من أنها تنهض بأدوار حاسمة في تكوين العنكبوتية الدلالية.

وكما يتضح من الشكل رقم ١٢/ ١، فإن الدلالة التي يتم تجميعها وتنظيمها ونشرها عن طريق العنكبوتية الدلالية، يتم اشتقاقها، وكذلك تحديدها أو تعريفها بالطبقات الثلاث الخاصة بهندستها المعمارية، التي نحاول استكشافها في هذا السياق. وتتكون الطبقة الثانية من ثلاثة عناصر، هي لغة الترميز القابلة للامتداد، التي تكفل القدرة على إنشاء ما وراء البيانات الخاصة بمصادر العنكبوتية، فيما يخرج عن حدود مستوى النظْم، بينما ييسر كل من فضاء الاسم **Namerpace (NS)**، وخطة لغة الترميز القابلة للامتداد **xmlchema** (أي إطار ترميز لغة الترميز القابلة للامتداد) استخدام قابلية ما وراء البيانات للتشغيل التبادلي، في لغة الترميز القابلة للامتداد **XML**. وعلى عكس لغة ترميز النصوص الفائقة **Hyper Text Markup Language (HTML)**، التي تصوغ **format** ما هو موجود على العنكبوتية فقط (كالبنط والحجم على سبيل المثال) يمكن للغة الترميز القابلة للامتداد أن تصوغ معلومات العنكبوتية، بنفس طريقة تكشيف قاعدة البيانات بالحقل (كالمؤلف والعنوان على سبيل المثال). فهب أن لدينا المؤلف **John Doe**، الذي نشر المقالة **B**. وباستخدام لغة ترميز النصوص الفائقة، فإن كل ما نستطيع عمله هو صياغة **John Doe** و**Article B**، بخيارات الصياغة المتاحة، كالحروف المائلة **italics** وبنط التايمز الروماني الجديد **Times New Roman**، وحجم الخط ١٢، على سبيل المثال.

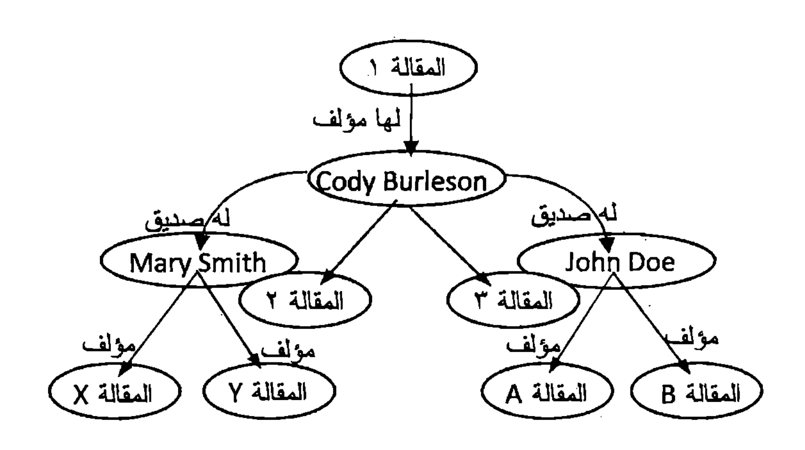
غير أنه في حالة لغة الترميز القابلة للامتداد **XML**، يمكن ترميز المعلومات نفسها على شكل: **<author> John Doe </author>** و **<publication> article B </publication>** على التوالي. ونظرا لأن ترميز لغة الترميز القابلة

للامتداد، يعمل بطريقة مماثلة إلى حد بعيد للتكشيف بالحقل، أحد خطوات إنشاء قواعد البيانات التقليدية؛ فإنه ينبغي توفير تقنية ما للتحكم، لضمان قابلية الترميزات الناشئة، بمختلف النظم، في مجالات مختلفة، للتبادل أو التشغيل التبادلي فيما بعد؛ فـ **<publication>...</publication>** على سبيل المثال، هو ترميز لغة الترميز القابلة للامتداد بالنسبة لجميع أنواع النصوص، ولا ينبغي استعمال شكل آخر **‹Document›...‹/Document›** على سبيل المثال. وفضاء الاسم **NS** وخطة لغة الترميز القابلة للامتداد **xmlschema** تقنيتان للتحكم، لمثل هذه الأغراض. ويمثل فضاء الاسم **NS** أحد المجالات (كالتجارة الإلكترونية المحلية على سبيل المثال)، وتوفر خطة لغة الترميز القابلة للامتداد، قواعد وتوجيهات لتكوين ترميزات هذه اللغة بالنسبة للمجال المستهدف.

ويستخدم إطار وصف المصادر (**RDF**، راجع أيضا ٣/ ١/ ٣/ ٢)، وخطة إطار وصف المصادر **rdfschema**، بوصفهما الطبقة الثالثة في الهندسة المعمارية للعنكبوتية الدلالية، للتعبير عن العلاقات فيما بين ما وراء البيانات (أي ترميزات لغة الترميز القابلة للامتداد) التي تعد لمعلومات العنكبوتية، في الأشكال المماثلة لإطار وصف المصادر (الشكل رقم ١٢/ ٢). وفي حالة "المؤلف **John Doe** الذي نشر المقالة **B**"، على سبيل المثال، فإن العلاقة في هذه العبارة (أي **John Doe** هو "مؤلف المقالة **B**"، موضحة في الركن الأسفل الأيمن في الشكل رقم ١٢/ ٢. ووظيفة خطة إطار وصف المصادر،كوظيفة خطة لغة الترميز القابلة للامتداد، هي ضمان أن تكون جميع الأشكال البيانية لإطار وصف المصادر قد نشأت باتباع خطة مماثلة، بحيث تكون هذه الرسوم البيانية قابلة للتشغيل التبادلي بوجه عام، في العنكبوتية الدلالية.



الشكل رقم ١٢/ ١ الهندسة المعمارية للعنكبوتية الدلالية



الشكل رقم ١٢/ ٢ مثال لرسم بياني لإطار وصف المصادر

وكما بينا في ٤/ ٤/ ٣، فإن التصنيف التخصصي عنصر جوهري في العنكبوتية الدلالية؛ نظرا لأن نجاح العنكبوتية الدلالية العالمية يتوقف في النهاية على هذا العنصر، لتحقيق التعامل فيما بين مصادر العنكبوتية، التي يتم ترميزها بلغة الترميز القابلة للامتداد، والتعبير عنها بالرسوم البيانية لإطار وصف المصادر، في جميع أنحاء العالم، إلى جانب القواعد المنطقية التي تقابلها بالنسبة للبرهان. ويمكن النظر إلي التصنيف التخصصي، على نحو ما، بوصفه لغة مقيدة عامة (كخطط التصنيف أو المكانز على سبيل المثال) بالنسبة لتنظيم مصادر العنكبوتية. ووجه الاختلاف الرئيس بين التصنيف التخصصي **ontology** واللغة المقيدة التقليدية، هو أن الأول قد قصد به التعامل مع العنكبوتية برمتها، بوسائل آلية، بينما الآخر يدوي بوجه عام، ولا يطبق مطلقا على نطاق يغطي جميع المجالات في العالم. وبعبارة أخرى، فإن لغة التصنيف التخصصي، الطبقة الرابعة في الهندسة المعمارية للعنكبوتية الدلالية، يمكن أن تساعد الحاسب في فهم معنى مصادر العنكبوتية، التي يتم تجهيزها في الطبقتين السابقتين، عندما يستخدم ذلك التصنيف على نطاق عام.

وتتضح نقطتان من هذا الوصف للطبقات الثانية والثالثة والرابعة للهندسة المعمارية للعنكبوتية الدلالية؛ أولاهما أن ما يمكن للعنكبوتية الدلالية القيام به، من حيث استنباط دلالات المعلومات، يمكن أن يبدو مثيرا وطموحا في الوقت نفسه. وثانيتهما أن كل ما يمكن أن تنهض به الطبقات الثلاث، ليس أكثر من تنظيم المعلومات على منصة العنكبوتية، ويتراوح هذا التنظيم بين التكشيف بالحقل (أي ترميز لغة الترميز القابلة للامتداد) والضبط الشامل للغة بالنسبة للعملية (كلغة التصنيف التخصصي على سبيل المثال).

العنكبوتية الدلالية وتنظيم المعلومات واسترجاعها:

وفقا للهندسة المعمارية للعنكبوتية الدلالية، التي عرضنا لها في ١٢/ ٣/ ١، فإن للعنكبوتية الدلالية آثارها واسعة المدى، بالنسبة لتنظيم المعلومات واسترجاعها. والتنظيم والاسترجاع الدلالي للمعلومات على العنكبوتية، والوكلاء الأذكياء، مثالان لمثل هذه الآثار.

التنظيم والاسترجاع الدلالي للمعلومات على العنكبوتية:

يتسم استرجاع معلومات العنكبوتية بالسمعة السيئة؛ نظرا لتدني مستوى الدقة أو التحقيق وارتفاع التشويش. وبينما نجد هناك كثيرا من العوامل (كغياب الضبط أو التحكم في المدخلات على سبيل المثال)، التي تسفر عن هذه المشكلة، يظل تكشيف الكلمات المفتاحية لمعلومات العنكبوتية، السبب الرئيس. وقد بذلت جهود مختلفة (كاستعمال رؤوس الموضوعات للتكشيف على سبيل المثال) لتوفير الكلمات غير المفتاحية، أو بمزيد من الدقة، التنظيم المفاهيمي لمعلومات العنكبوتية. إلا أن هذه الجهود قد بذلت على نطاق محدود فقط، ولم يكن لها تأثير يذكر، على الارتفاع بمستوى أداء تنظيم المعلومات واسترجاعها في العنكبوتية، في ظل وجود الكميات الهائلة من المعلومات في هذه البيئة، بالإضافة إلى الخصائص الأخرى لمعلومات العنكبوتية.

وتتعلق قضية التحقيق المنخفض والتشويش المرتفع، في استرجاع معلومات العنكبوتية، في الأساس، بالدلالات أو "المعاني المقصودة المستعملة في التواصل البشري"، كما عبر عن ذلك بيرنرز-لي، وهول، وهندلر **Berners-Lee, Hall & Hendler (2006, P. 770)**. فإذا كان بإمكان العنكبوتية الدلالية استنباط الدلالات من معلومات العنكبوتية، كما كان من المنتظر، لأمكن لمشكلة التحقيق المنخفض التي يعاني منها استرجاع المعلومات في العنكبوتية، أن تحل، أو يتم التخفيف من حدتها على الأقل؛ فاستعمال لغة الترميز القابلة للامتداد **XML**، على سبيل المثال، في ترميز

معلومات العنكبوتية، يمكن أن يكفل للمستفيد أن يقول وداعا لاسترجاع معلومات العنكبوتية غير محددة الحقول، ويرحب باستعمال البحث بالحقل. وفضلا عن ذلك، فإن التحكم في اللغة، في تنظيم المعلومات واسترجاعها في العنكبوتية، يمكن أن يكفل للمستفيد القدرة على التحقق، على سبيل المثال، مما إذا كانت الكلمة **bank** تعني مؤسسة مالية، أم الأرض المرتفعة على ضفتي النهر. ويمكن من ناحية أخرى، لتنظيم المعلومات الدلالية واسترجاعها، أن يكون بإيجاز، حقيقة واقعة بمجرد أن تتطور العنكبوتية الدلالية.

الوكلاء الأذكياء:

بإمكان الوكلاء الأذكياء، باستخدام أساليب ذكاء اصطناعي متنوعة، وفي شكل برمجيات الحاسبات، إنجاز مهام معينة دون تدخل بشري. ومن ناحية أخرى فإن النظم الخبيرة تعرف بأنها برمجيات حاسبية تدلي بالمعلومات، أو تصدر التوصيات، أو تحل المشكلات، بطريقة وعلى مستوى أداء، يضاهي ذلك الذي يبديه الخبير البشري في المجال. والنظم الخبيرة في جوهرها وكلاء أذكياء على نطاق واسع، أو أسلاف الوكلاء الأذكياء. ويتزايد استعمال كلمة بوت **bots** (الشكل المختصر لكلمة **robots**) بوصفها مرادفة للوكلاء الأذكياء.

وقد انتشر مفهوم الوكلاء الأذكياء، إلى حد ما، بفضل بيرنرز-لي، وهندلر، ولاسيلا **Berners-Lee, Hendler & Lassila (2001)**، عندما قدموا رسميا، العنكبوتية الدلالية للجمهور العام؛ ففي تلك المقالة(*)، وصف المؤلفون الثلاثة بوضوح، وكيلا قادرا على جدولة المواعيد الطبية. وقد استعرض دنت **Dent (2007)**

(*) تقصد المقالة التي تناول فيها المؤلفون الثلاثة فكرة العنكبوتية الدلالية. وقد وردت بيانات تلك المقالة كاملة في قائمة مراجع هذا الفصل. (المترجم)

كثيرا من الجوانب الخاصة بالوكلاء الأذكياء، كخصائصهم ووظائفهم على سبيل المثال، من منظور المكتبات الحديثة. وكما ذهب هندلر **(1999) Hendler** فإن الوكيل الذكي المثالي، ينبغي أن يتمتع بأربع خصائص:

١. تواصليا: القدرة على فهم أهداف المستفيد وأفضلياته وحدوده.

٢. قادرًا: أي القدرة على حسم الخيارات، لا مجرد تقديم النصح ببساطة.

٣. مسيَّرا ذاتيا: القدرة على التصرف دون أن يكون تحت سيطرة المستفيد طوال الوقت.

٤. تكيفيا: القدرة على التعلم من الخبرة المكتسبة، حول مهامه وأفضليات المستفيد منه.

وتطبيقات الوكلاء الأذكياء كثيرة ومتنوعة، في نطاق مجال تنظيم المعلومات واسترجاعها، وخارج هذا المجال على السواء. وأيا كان الهدف الذي من أجله يتم تصميم الوكيل الذكي، فإنه ينبغي أن تتوافر الخصائص الأربع التي حددها هندلر. ويمكن الإفادة من الوكلاء الأذكياء في كثير من المواقف المختلفة، التي تتراوح بين أمور حياتنا اليومية، ومشروعات البحوث والتطوير المعقدة. ومن الممكن العثور على عدد كبير من الوكلاء الأذكياء، وخصوصا على الصعيد العام، في موقع بوتسبوت **.Botspot (www.botspot.com)**

ويستهدف الوكلاء الأذكياء، بالنسبة لتنظيم المعلومات واسترجاعها، فروعا محددة بعينها في المجال؛ فمهام تنظيم المعلومات واسترجاعها، المعروفة على أحسن وجه، بشكل عام، يمكن التحقق منها صراحة، ولها حدود واضحة المعالم، يمكن للوكلاء الأذكياء النهوض بها الآن. والبث الانتقائي للمعلومات **(SD1)** أحد أمثلة هذه المهام، إذ يمكن للوكلاء الأذكياء مراقبة نظام واحد لاسترجاع المعلومات أو

أكثر، وإخطار المستفيد عندما تـرد إلى النظـام معلومـات جديـدة في موضـوع معـين. والتقسيم الفئوي للنتـائج المسترجعة، المألوف في نظـم اسـترجاع الإنترنـت، نظـرا لضخامة كم مخرجات هذه النظم، مثال آخر في هذا الـصدد. وأشـهر أمثلـة الـوكلاء الأذكياء في تنظيم المعلومات واسترجاعها، بـرامج تجميـع البيانـات لـنظم اسـترجاع معلومات الإنترنت، وتـسمى أحيانـا بـالزواحف **crawlers**، أو الـشواخص الآليـة **robots**، أو العناكب. ويزور هؤلاء الوكلاء المواقع المستهدفة بانتظام، ويستنـسخون معلومات معينة، ويرسلونها إلى قاعدة بيانات النظام.

ولا يمكن لتحقيق التكامل بين المعلومات المسترجعة مـن مـصادر مختلفـة، أن يكون مهمة عصية بالنسبة للوكلاء الأذكياء. ويشكل الوكلاء الأذكياء بالنـسبة لمهـام تنظيم المعلومات واسترجاعها، التي تفرض تحديات أكثر، كـصياغة الاستفـسارات، وتمثيل المعلومات أو التعبير عنها، موضوعات للبحث في الوقت الراهن. فقـد قـصد على سبيل المثال، بأساليب الرد على الاستفسارات، التي تطبق في بعض نظم استرجاع المعلومات، التعرف عن طريق توجيه الأسئلة، على مـا يريـده المـستفيد فعـلا، عنـدما يكون بصدد إعداد طلب للمعلومـات. ومحـاولات تطـوير وكـلاء أذكيـاء بإمكـانهم تجميع الملخصات والمستخلصات الآلية، تجري الآن أيضا على قدم وساق، وخصوصا بالنـسبة للمعلومـات الإخباريـة، ومعلومـات إدارة الأعـمال (راجـع ١٢/ ٢/ ٢). ويتوقف نجاح هذه الجهود في النهاية، على البحث في الذكاء الاصطناعي.

ومن الممكن النظر إلى إضفاء الطابع الشخصي عـلى عمليـات البحـث، بوجـه عام، بوصفه امتدادا للوكلاء الأذكياء، حيث يكون نظام استرجاع المعلومات مـصمما على نحو يتيح إمكانية تجميع سمات المستفيد **user profile**، وتاريخ عمليـة البحـث، وغيرها من المعلومات المتصلة بذلك المـستفيد، معـا لتـشكيل تقنيـة أشـبه بالوكيـل. ويمكن تنشيط هذه التقنية في عملية الاسترجاع، لمساعدة نظام استرجاع المعلومات في

التعرف على ما يريده المستفيد فعلا في الموقف الراهن. فالمستفيد على سبيل المثال، قد يُدخل كلمة **Java** كاستفسار في صندوق البحث، وحينئذ تقوم تقنية البحث ذي الطابع الشخصي فورًا، بإجراء سلسلة من المهام (كمراجعة سمات المستفيد وتاريخ عملية البحث، على سبيل المثال) للكشف عما قصد بمعنى **Java** في هذه الحالة. وإذا كانت هناك للمستفيد خلفية في البرمجة، فإنه من المرجح أن تكون **Java** لغة للبرمجة، يود المستفيد العثور عليها. وإذا كانت سمات المستفيد لا تشتمل على معلومات حول البرمجة، إلا أنه يتضح من تاريخ عملية البحث الخاصة به، أن مواقع وكالات السفر قد تمت زيارتها مؤخرا، فإن **Java** يمكن أن تدل هنا على جزيزة جافا **Java** في إندونيسيا. كذلك يمكن التحقق مما إذا كانت **Java** نوعا من البن أم لها معانٍ أخرى، على النحو نفسه، في نظام لاسترجاع المعلومات يدعم البحث الشخصي. وتخرج قضية الخصوصية المتصلة بالبحث الشخصي عن نطاق اهتمام هذا الكتاب.

التحديات التي تواجهها العنكبوتية الدلالية كأحد تطبيقات الذكاء الاصطناعي:

على الرغم من التقدم الملحوظ في تطور العنكبوتية الدلالية، يتساءل كثير من الباحثين (من أمثال **Legg, 2007; O'Hara, 2004; Spärck Jones, 2004**) حول جدوى هذا التطبيق الخاص بالذكاء الاصطناعي. والعنكبوتية الدلالية، لم يقصد بها أن يتم إنشاؤها وصيانتها عن طريق البشر، كما كان يحدث في تنظيم المعلومات واسترجاعها بالأساليب التقليدية، وإنما يمكن استخدام تقنيات الذكاء الاصطناعي بكثافة، في تطوير العنكبوتية الدلالية، وما يتبع ذلك من عمليات، على النحو الذي يمكن أن يثير أسئلة أكثر مما يمكن للباحثين في المجال الإجابة عليه. أما فكرة الدلالة التي تمثل أساس هذا التطبيق، فتضغط من ناحية أخرى على زناد المزيد من التساؤلات؛ نظرا لأن كيفية إنجاز البحث المفاهيمي **concept searching** في بيئة آلية، كانت منذ زمن بعيد، مشكلة جوهرية تحظى بالدرس من جانب كثيرين في المجال.

وهناك في المقام الأول، نوعان من التحديات التي تواجه رؤية العنكبوتية الدلالية؛ يتعلق أولهما بهذه العنكبوتية بوصفها أحد تطبيقات الذكاء الاصطناعي، ويتصل الآخر بقدرتها المنتظرة على نقل الدلالات بالوسائل الآلية. ونتناول النوع الأول من التحديات في ١٢/ ٤، عندما نلقي نظرة فاحصة على العلاقة بين الذكاء الاصطناعي وتنظيم المعلومات واسترجاعها. أما الضرب الآخر من التحديات فيمكن تلخيصه فيما يلي:

أولا، من الذي يمكن أن يتحمل مسئولية ترميز جميع مصادر معلومات العنكبوتية بلغة الترميز القابلة للامتداد XML، ثم التعبير عنها بالرسوم البيانية بأشكال إطار وصف المصادر RDF؟ ثانيا، من الذي يمكن أن يتحمل مسئولية تطوير وصيانة خطط لغة الترميز القابلة للامتداد **xmlschemas**، وخطط إطار وصف المصادر **rdfrchemas**، وخطط التصنيف التخصصية **ontologies** اللازمة لضمان استعمال الدلالات المشتركة، الناشئة عن الهندسة المعمارية للعنكبوتية الدلالية، والتشغيل التبادلي لهذه الدلالات؟ ثالثا، هل يمكن وضع تصنيف واحد، يمكن أن يطبق في جميع المجالات واللغات والثقافات الموجودة على هذا الكوكب؟ ويتم طرح هذه التساؤلات الثلاثة في الأساس، من جانب الباحثين من خارج تجمع العنكبوتية العالمية **W3C** (من أمثال **Legg, 2007; O'Hara, 2004; Spärck Jones, 2004**). إلا أن بيرنرز-لي ورفاقه أنفسهم (مثل **Berners-Lee, Hall & Hendler, 2006**) يطرحون تساؤلات أخرى. ومن بين التساؤلات التي تطرحها هذه المجموعة، على سبيل المثال، كيف يمكن جعل كل المعلومات التي تتم معالجتها للعنكبوتية الدلالية، قابلة للتداول في المجال العام، نظرا لأن التعامل المجاني **open access** أحد الشروط الأساس التي ينبغي للعنكبوتية الدلالية الوفاء بها.

وقد أثار شادبولت، وهول، وبيرنـرز-لي Shadbolt, Hall and Berners-Lee (2006) المزيد من التساؤلات المتعلقة بالعنكبوتية الدلاليـة بوجـه عـام، عنـدما عاودوا النظر في هذه الرؤية؛ كيـف عـلى سـبيل المثـال، يمكننـا اسـتجواب الأعـداد الضخمة من مستودعات المعلومات اللامركزية، متفاوتة الأحجام، على نحو فعـال؟ كيف ننـسق، ونـضع خـرائط الـربط بـين خطـط التـصنيف؟ كيـف نبنـي متـصفحاً للعنكبوتية الدلالية، لتصور الرسوم البيانية لإطار وصف المصادر، المترابطة الضخمة، والملاحة فيها بشكل فعال؟ كيف نبني الثقة والمصداقية في المحتوى؟

ووجهات النظر في العنكبوتية الدلالية في تغير مستمر، نتيجة للبحوث الجارية، والتطورات التقنية؛ فالتصنيف، على سبيل المثال، كان ينظر إليه بوصـفه كيانـا واحـدا يمكن أن يشمل معارف العالم بأسره، في الهندسة المعمارية للعنكبوتية الدلالية، الواردة في الشكل رقم ١٢/ ١. كما تبين فيما بعد، أن وضع خطط التـصنيف التـي تركـز عـلى قطاعات صغيرة، أو على مجالات واضحة المعالم والحدود، ليس بـالأمر غـير الـواقعي فحسب، وإنما لا يمكن تطبيقه أيضا، على النحو الـذي كـان يـتم بـه إنـشاء القواعـد المعرفية للنظم الذكية في الأيام الخوالي. ومن الصعب عادة التنبؤ بالمـستقبل. ويـصدق الأمر نفسه، وعلى نحو أكثر تأكيدا، فيما يتعلق بمستقبل العنكبوتية الدلالية.

الذكاء الاصطناعي وتنظيم المعلومات واسترجاعها:

إلى أي مـدى يمكـن للـذكاء الاصـطناعي المـضي قـدما في تنظـيم المعلومـات واسترجاعها؟ ولا نجد إجابة لهذا التساؤل حتى الآن؛ نظرا لأنه في المقام الأول، يبدو بناء وبرمجة تقنية يمكن أن تحاكي بنجاح الأنشطة الذكية للبشر، ليس بالمهمة البسيطة. ويمكـن لـذلك أن يكـون أصـعب حبـة جـوز يمكـن كـسرها، في بحـوث الـذكاء الاصطناعي (Perscovitz, 1999).

ومن ثم فإن دور الذكاء الاصطناعي في تنظيم المعلومات واسترجاعها اليوم محدود، ولا ينبغي المبالغة في هذا الدور، لسببين؛ أولهما أن المهام المعقدة التي ينطوي عليها تنظيم المعلومات واسترجاعها، تعتمد على رصيد معرفي واسع المدى، كما تعتمد على استخدام هذا الرصيد على نحو مرن وبناء. وثانيهما أن كثيرا من عمليات تنظيم المعلومات واسترجاعها، لا تتم إلا على نحو سطحي ضحل، حتى وإن كان ذلك يتم بواسطة البشر (**Spärck Jones, 1991**). وقد علق سوانسون **Swanson (1988)** ذات مرة قائلا "على الرغم من أن هدف كثير من البحوث المتقدمة في استرجاع المعلومات، هو جعل الحاسبات كيانات أشبه بالبشر، فإن احتمال أن يكون الناس أشبه بالبشر لا ينبغي التغاضي عنه".

ولا ينبغي فعلا التغاضي عن دور البشر في تطوير تطبيقات الذكاء الاصطناعي، لأغراض تنظيم المعلومات واسترجاعها. فالحاسب لا يتمتع بحدة الذهن، أو العمق، أو الاتساع، أو الثراء الذي يتسم به الذكاء البشري؛ نظرا لأن الحاسب لا يزال أبسط مليون مرة من العقل البشري (**Hefner, 2000**). إلا أنه بإمكان الذكاء الاصطناعي، في وضعه الراهن، أن يقدم إسهاما، في المواقف التي يمكن أن تكون فيها القاعدة المعرفية أو الاستدلال، محدودين في جوهرهما، أو يكون بإمكان المستفيد تعويض مظاهر القصور. ويمكن أيضا تطبيق الذكاء الاصطناعي في الظروف التي تستخدم فيها خوارزميات هذا الذكاء إلى جانب الأساليب الأخرى (**Spärck Jones, 1991**).

والتطبيقات الخاصة بتنظيم المعلومات واسترجاعها، المعتمدة على الذكاء الاصطناعي، وحده أو بالإضافة إلى الأساليب الأخرى في نمو، إلا أن كل أمثال هذه التطبيقات تقريبا، إما أن تكون ما تزال في مختبرات البحث، أو تكون قد اكتسبت فعلا الطابع التجاري، وتتاح مقابل رسوم. ومن ناحية أخرى، تم إدخال مختلف عناصر

الذكاء الاصطناعي في كثير من التطبيقات الخاصة بتنظيم المعلومات واسترجاعها. وقد تناول ولكس (2005) Wilks بالتفصيل، بعض تجارب تطبيقات الذكاء الاصطناعي في مجال تنظيم المعلومات واسترجاعها، وذلك في سياق إلقاء نظرة فاحصة على العلاقة بين الذكاء الاصطناعي واسترجاع المعلومات. ففي أي اتجاه إذن يسير مجال تنظيم المعلومات واسترجاعها، ونحن على مشارف العصر الرقمي؟ ويوحي البحث والتطوير في الذكاء الاصطناعي ببعض الاحتمالات على الأقل.

المراجع

Berners-Lee, Tim. (2000). Semantic web on XML (Slide 10). [Presentation slides]. Retrieved January 18, 2009, from www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html

Berners-Lee, Tim, Hall, Wendy, and Hendler, James. (2006). Creating a science of the web. *Science,* 313(5788), 769-771.

Berners-Lee, Tim, Hendler, James, and Lassila, Ora. (2001). The semantic web. *Scientific American,* 284(5), 35-43.

Campbell, Philip. (2008). Community cleverness needed. [Editorial]. *Nature,* 455(7209), 1.

Carbonell, Jaime, et al. (2000). Vision statement to guide research in question answering (QA) and text summarization. Retrieved January 14,2009, from www-nlpir.nist.gov/projects/duc/papers/Final-Vision-Paper-v1a.pdf

Cater, Arthur W. S. (2005). Question answering. In John I. Tait (Ed.), *Charting a new course: Natural language processing and information retrieval* (pp. 105-128). New York: Springer.

Cercone, Nick, and McCalla, Gordon. (1984). Artificial intelligence: Underlying assumptions and basic objectives. *Journal of the American Society for Information Science,* 35(5), 280-290.

Chen, Hsinchun. (1995). Machine learning for information retrieval: Neutral networks, symbolic learning, and genetic algorithms. *Journal of the American Society for Information Science,* 46(3), 194-216.

Chen, Hsinchun, et al. (1994). Automatic concept classification of text from electronic meetings. *Communications of the ACM,* 37(10), 56-73.

Chowdhury, Gobinda G. (1999). *Introduction to modern information retrieval.* London: Library Association Publishing.

Dent, Valeda F. (2007). Intelligent agent concepts in the modern library. *Library Hi Tech*, 25(1), 108-125.

Doszkocs, T. E. (1986). Natural language processing in information retrieval. *Journal of the American Society for Information Science*, 37(4), 191-196.

El Hadi, Widad Mustafa. (2003). Human language technology and its role in information access and management. *Cataloging & Classification Quarterly*, 37(1/2), 131-151.

Flesca, Sergio, et al. (2004). Web wrapper induction: A brief survey. *AI Communications*, 17(2), 56-61.

Godby, C.Jean,and Reighart, Ray R. (2001). The Word Smith toolkit. *Journal of Library Administration*, 34(3/4), 307-316.

Greenberg, Jane. (2003). Metadata: The fundamental component of the semantic web. *Bulletin of the American Society for Information Science and Technology*, 29(4), 16-18.

Gregg, Dawn G., and Walczak, Steven. (2006). Adaptive web information extraction. *Communications of the ACM*, 49(5), 78-84.

Hefner, Katie. (December 28,2000). Artificial intelligence hasn't peaked (yet). *New York Times*. Retrieved October 12, 2009, from www.nytimes.com/ 2000/12/28/technology/28ARTI.html

Hendler, James. (March 11,1999). Is there an intelligent agent in your future? *Nature*, 398(2763). Retrieved January 21, 2009, from www.nature.com/ nature/webmatters/agents/agents.html

Jennings, Andrew, and Higuchi, Hudeyuki. (1992). A browser with a neural network user model. *Library Hi Tech*, 10(1/2), 77-93.

Kwok, Cody, Etzioni, Oren, and Weld, Daniel S. (2001). Scaling question answering to the web. *ACM Transactions on Information Systems*, 19(3), 242-262.

Lau, Annie Y. S., and Coiera, Enrico W. (2006). A Bayesian model that predicts the impact of web searching on decision making. *Journal of the American Society for Information Science and Technology*, 57(7), 873-880.

Legg, Catherine. (2007). Ontologies on the semantic web. *Annual Review of Information Science and Technology*, 41, 407-451.

Liddy, Elizabeth D. (1998). Enhanced text retrieval using natural language processing. *Bulletin of the American Society for Information Science and Technology,* 24(4), 14-16.

Lin, Jimmy, and Katz, Boris. (2006). Building a reusable test collection for question answering. *Journal of American Society for Information Science and Technology,* 57(7), 851-861.

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development,* 2(2), 159-165.

Mani, Inderjeet. (2001). *Automatic summarisation.* Amsterdam: John Benjamins.

Maybury, Mark T. (2002). *Toward a question answering roadmap.* Retrieved January 14, 2009, from www.mitre.org/work/tech_papers/tech_papers_02/maybury_toward/maybury_toward_qa.pdf

Maybury, Mark T. (2004). *New directions in question answering.* Cambridge, MA: MIT Press.

McDonald, Daniel M., and Chen, Hsinchun. (2006). Summary in context: Searching versus browsing. *ACM Transactions on Information Systems,* 24(1), 111-141.

Mojsillovic, Aleksandra, Gomes, Jose, and Rogowitz, Bernice. (2004). Semantic-friendly indexing and querying of images based on the extraction of the objective semantic cues. *International Journal of Computer Vision,* 56(1/2), 79-107.

Mullins, Justin. (2005). Whatever happened to machines that think? *New Scientist,* 186(2496), 32-37.

Neumann, Gunter, and Xu, Feiyu. (2004). Mining natural language answers from the web. *Web Intelligence & Agent Systems,* 2(2), 123-235.

O'Hara, Kieron. (2004). Ontologies and technologies: Knowledge representation or misrepresentation. *ACM SIGIR Forum,* 38(2), 11-17.

Over, Paul, Dang, Hoa, and Harman, Donna. (2007). DUC in context. *Information Processing & Management,* 43(6), 1506-1520.

Pescovitz, David. (March 18,1999). Look what's talking: Software robots. *New York Times.* Retrieved October 12, 2009, from www.nytimes.com/library/ tech/99/03/circuits/articles/18bots.html

Pomerantz, Jeffrey. (2005). A linguistic analysis of question taxonomies. *Journal of the American Society for Information Science and Technology,* 56(7), 715-728.

Shadbolt, Nigel, Hall, Wendy, and Berners-Lee, Tim. (2006). The semantic web revisited. *IEEE Intelligent Systems,* 21(3), 96-101.

Spärck Jones, Karen. (1991). The role of artificial intelligence in information retrieval. *Journal of the American Society for Information Science,* 42(8), 558-565.

Spärck Jones, Karen. (2004). What's new about the semantic web? Some questions. *ACM SIGIR Forum,* 38(2), 18-23.

Spärck Jones, Karen. (2007). Automatic summarising: The state of the art. *Information Processing & Management,* 43(6), 1449-1481.

Swanson, Don R. (1988). Historical note: Information retrieval and the future of an illusion. *Journal of the American Society for Information Science,* 39(2), 92-98.

Turmo, Jordi, Ageno, Alicia, and Gatala, Neus. (2006). Adaptive information extraction. *ACM Computing Surveys,* 38 (2), 1-47.

Voorhees, Ellen M. (2005). Question answering in TREC. In Ellen M. Voorhees and Donna K. Harman (Eds.), *TREC: Experiment and evaluation in information retrieval* (pp. 233-257). Cambridge, MA: MIT Press.

Wilks, Yorick. (2005). Unhappy bedfellows: The relationship of AI and IR. In John I. Tait (Ed.), *Charting a new course: Natural language processing and information retrieval* (pp. 255-282). New York: Springer.

ثبت المصطلحات

(A)

المحتوى / الموضوع	Aboutness
المستخلصات	Abstracts
نقاط الوصول	Access points
البحث المخصص/ الموجه	Ad hoc searching
البحث بالتجاور	Adjacency searching
الخوارزميات (نسبة إلى الخوارزمي، عالم الرياضيات)	Algorithms
التناظري	Analog
قواعد الفهرسة الأنجلو-أمريكية	Anglo - American Cataloging Rules (AACR)
الحالة المعرفية غير السوية	Anomalies state of knowledge (ASK)
الذكاء الاصطناعي	Artificial intelligence (AI)
التكشيف بالتعيين / التخصيص	Assignment indexing
التفكير الترابطي / التداعي	Associative thinking
المعلومات الصوتية / السمعية	Audio information
التكشيف بمساعدة الآلات	Automated indexing
الترجمة الآلية	Automatic translation

(B)

Backlinks	الروابط الراجعة
Batch mode	التعامل على دفعات
Bayesian inference	الاستدلال الافتراضي
Bibliographic coupling	المزاوجة الوراقية
Bibliometrics	القياسات الوراقية
Blogging	التدوين
Book reviews	مراجعات الكتب
Boolean searching	البحث البوليني
Buzzwords	الكلمات / العبارات الطنانة

(C)

Case - sensitive searching	البحث بمراعاة الحرف الأول
Case sensitivity	مراعاة الحرف الأول من الكلمة
Casual Browsing	التصفح العارض
Categorization	التصنيف / التقسيم الفئوي
Chain indexing	التكشيف التسلسلي
Citation	الاستشهاد المرجعي
Clustering	التجمع العنقود
Co-citation	المصاحبة الوراقية
Cognition	المعرفة
Command language	لغة الأوامر
Communication	التواصل/ الاتصال
Compact Disk-Read Only Memory	الأسطوانات المكتنزة

Compatibility	التوافق / التناغم
Concept analysis	تحليل المفاهيم / الموضوعات
Concept coordination	ربط المفاهيم / الموضوعات
Concept searching	البحث بالمفاهيم
Concordance	كشاف النصوص
Conflation	الخلـط
Consistency	الاطـراد
Controlled language -Controlled vocabulary	اللغة المقيدة / المقننة
Coordinate indexing	التكشيف الترابطي
Cost-benefit	عائد التكلفة
Cost-effectiveness	فاعلية التكلفة
Cranfield Tests	تجارب كرانفيلد
Crawlers	الزواحـف
Critical abstract	المستخلص النقدي
Current Awareness	الإطاحة الجاريـة

(D)

Database	قاعدة البيانات/ مرصد البيانات
Data mining	التنقيب عن المعلومات
Derivative indexing	التكشيف الاشتقاقي
Descriptors	الواصفات
Digital	الرقمـي
Digital Object Identifiers (DOI)	محددات هوية الكيانات الرقمية

Digital Versatile Disk (DVD)	الأسطوانة الرقمية متعددة الأغراض
Directed Browsing	التصفح الموجه
Dublen Core	بـؤرة دبلـن

(E)

Effectiveness	الفاعليـة
Enumeration	الحصـر
Enumerative	الحصـري
Exact matching	المضاهاة التامة
Expert Systems	النظم الخبيرة
Exploratory browsing	التصفح الاستكشافي
Extensible Markup Language (XML)	لغة الترميز القابلة للتوسع

(F)

Fallout	التسرب
False coordination	الربط المزيف
Federal searching	البحث في مراصد بيانات متعددة/ البحث الموحد
Feedback	التلقيـم المرتـد
Field	الحقـل
Field search	البحث بالحقـل
File	الملـف

File Transfer Protocol	بروتوكول نقل الملفات
Folksonomy	التصنيف المتعارف عليه
Function Words	الكلمات الوظيفية / النحوية
Fuzzy searching	البحث المشوش

(G)

Gateway	البوابـة
Graphical	التصويري
Go List	قائمة الاعتبار

(H)

Hits	التسديدات
Homepage	الصفحة الرئيس
Hyperlinks	الروابط الفائقة
Hyperstructure	التنظيم الفائق
Hypertext Markup Language (HTML)	لغة ترميز النصوص الفائقة

(I)

Identification code	ترميز التحقق
Identifier	معرِّف / محدد الهوية
Index terms	المصطلحات الكشفية
Indexing	التكشيف
Infrastructure	البنية الأساس
Intelligent Agents	الوكـلاء الأذكياء

Internet Retrieval Systems	نظم استرجاع الإنترنت
Interoperability	التوافقية / القابلية للتشغيل التبادلي
Inverse Document Frequency (IDF)	التردد العكسي للوثائق
Inverted File	الملف المقلوب / المصنف
Indicative abstract	المستخلص الكشفي
Inductive learning	التعلم الاستقرائي
Inference	الاستدلال
Inference engine	محرك الاستدلال
Inference networks	شبكات الاستدلال
Infix	الحشو (في الصرف)
Information access	الوصول إلى المعلومات
Informatics	تقنيات المعلومات
Information representation	تمثيل / تنظيم المعلومات
Information retrieval	استرجاع المعلومات
Information searching	البحث عن المعلومات
Information seeking	التماس المعلومات / البحث عن المعلومات
Information storage	اختزان المعلومات
Information technology	تقنيات المعلومات
Informetrics	قياسات المعلومات

(J)

Juxtaposition Searching	البحث بالتجاور

(K)

Keywords	الكلمات المفتاحية
Keyword in Context Index (KWIC)	كشاف الكلمات المفتاحية في السياق
Keyword indexing	التكشيف بالكلمات المفتاحية
Keyword search	البحث بالكلمات المفتاحية
Knowledge base	القاعدة المعرفية / الأساس المعرفي

(L)

Literary warrant	المسوغ الأدبي / مسوغ الإنتاج الفكري
Locator	محدد المكان
Loud breath	التنفس الجهري

(M)

Machine indexing	التكشيف بالآلات
Machine Readable Cataloging	الفهرسة المقروءة آليا (مارك)MARC
Magnetic Resonance Imaging (MRI)	التصوير بالرنين المغناطيسي
Menu	قائمة الخيارات
Metadata	ما وراء البيانات
Metaretrieval systems	ما وراء نظم الاسترجاع
Metasearch engines	ما وراء محركات البحث
Middle truncation	بتر الحشو / بتر وسط الكلمة
Morpheme	المورفيم (أصغر وحدة صرفية)
Morphology	الصرف
Multimedia	الوسائط المتعددة

(N)

Navigation	الملاحـة
Negative searching	البحث السلبي
NEsted Phrase Indexing system (NEPHIS نفيز)	نظام تكشيف العبارات المسورة
Neural Networks	الشبكات العصبية
Newswires	الشبكات الإخبارية
Node	النقطة الارتكازية
Noise Factor	معامل (التشويش/ الصخب)
Notched Cards	البطاقات المثلومة

(O)

Objectivity	الموضوعية
Ofness	الوصف
Online Public Access Catalog (OAPC أوباك)	الفهرس المتاح على الخط المباشر
Online systems	نظم الخط المباشر
Ontology	التصنيف التخصصي
Open access	التعامل المجاني
Operator	المعامل
Optical Character Recognition (OCR)	التعرف على الأحرف ضوئيا
Organic User Interface (DUI)	واجهة المستفيد العضوية

(P)

Partial matching	المضاهاة الجزئية
Parsing	التحليل / الإعراب
Passage retrieval	استرجاع الفقرات
Pattern recognition	التعرف على الأنماط
Peer-to-peer (P2P) searching	بحث الانداد
Phrase searching	البحث بالعبارات
Pooling	التجميع البؤري
Portable	متعدد المنصات
Portal	المرفأ
Positional matching	المضاهاة الموضعية
Post-coordination	الربط اللاحق
Precision	التحقيق
Pre-coordination	الربط المسبق
PRE served Context Index System (PRECIS بريسيز)	نظام تكشيف السياق المصون
Proximity searching	البحث بالتقارب
Pull-down menu	قائمة الخيارات المنسدلة

(Q)

Quasi-Concept Searching	البحث شبه المفاهيمي / شبه الموضوعي
Quasi Fields	أشباه الحقول

Query by example	الاستفسار بالمثال
Query by Image Content (QBIC)	الاستفسار بمحتوى الصورة
Query Contraction	تقليص الاستفسار
Query expansion	توسعة الاستفسار
Query formulation	صياغة الاستفسار
Query modification	تعديل الاستفسار
Query parsing	تحليل الاستفسار
Question answering	الرد على الاستفسارات

(R)

Radio Frequency Identification (RFID)	التحقق بالتردد اللاسلكي
Range matching	المضاهاة بالمدى
Ranking	الترتيب الطبقي
Really Simple Syndication (RSS)	الإخطار البسيط فعلا
Recall	الاستدعاء
Record	التسجيلة
Reference interview	المقابلة المرجعية
Relative recall	الاستدعاء النسبي
Relevance	الصلاحية / الاتصال بالموضوع
Resource Description Framework (RDF)	إطار وصف المصادر
Reviews	المراجعات العلمية

Rich Site Summary (RSS)	ملخص الموقع الثري
Right truncation	بتر عجز الكلمة
Robots	الشخوص الآلية
Routing	التوجيه / تحديد المسار

(S)

Search engine	محرك البحث
Search mechanism	تقنية البحث
Search / Retrieve Web (SRW)	البحث والاسترجاع في العنكبوتية
Search strategy	استراتيجية البحث
Selective Dissemination of Information (SDI)	البث الانتقائي للمعلومات
Semantic image	الصورة الدلالية
Semantic networks	الشبكات الدلالية
Semantic Web	العنكبوتية الدلالية
Sequential file	الملف التسلسلي
Showcases	نوافذ العرض
Site	الموقـع
Speech recognition	التعرف على الحديث
Social Tagging	التمييز الاجتماعي
Spiders	العناكب
Spoken documents	الوثائق الناطقة
Stemming	رد الكلمة إلى الجذع

Stop list	قائمة الاستبعاد
String indexing	التكشيف التسلسلي
Stripping	تجريد الكلمة
Summarization	التلخيص
Surrogates	البدائـل
Symbolic reasoning	المنطق الرمزي
Syntax	النظْـم/ النحو
Synapses	نقاط الترابط العصبية
Synthesis	التركيب

(T)

Tag Cloud	تجمع التيجان
Tagging	التمييز
Tags	التيجـان
Taxonomy	التصنيف الهرمي
Techniques	تقنيات
Term Frequency	تردد المصطلحات
Terminals	المنافـذ
Term Masking	إخفاء معالم المصطلحات
Term Weighting	وزن المصطلحات

Threshold	الحد الأدنى / العتبة
Topic Sentences	الجمل الموضوعية
Truncation	البتـر

(U)

Unified Resource Identifier (URI)	المحدد الموحد لهوية المصدر
Unified Resource Locator (URL)	المحدد الموحد لمكان المصدر
Uniterm	المصطلح الأحادي
Usability	القابلية للاستخدام
User - friendliness	سهولة التعامل / التعاطف مع المستفيد
User - Friendly	سهولة التعامل / متعاطف مع المستفيد
User profile	سمات المستفيد
User warrant	مسوغ المستفيد
Utility	الجـدوى

(V)

Vector Space Model	أنموذج الفضاء المتجهي
Vocalization	التلفظ / النـطق
Vortals	المرافئ الرأسية / المتخصصة

(W)

Web 2.0	الجيل الثاني للعنكبوتية
Webmaster	مسئول العنكبوتية
Webpage	صفحة العنكبوتية
Weighte searching	البحث بالوزن
Wide Area Information Servers (WAIS)	ندل المعلومات واسعة المدى
Wildcard	التبديـل
World Wide Web (WWW)	العنكبوتية العالمية

المؤلفة في سطور:

هتنج تشو Heting Chu

ولدت في الصين، عام ١٩٥٧. وحصلت على ليسانس في علم المكتبات، في جامعة بكين، الصين، وعلى ماجستير المكتبات والمعلومات، في جامعة ماجل، كندا، وعلى دكتوراه دراسات المعلومات، في جامعة دركزل، الولايات المتحدة الأمريكية. وقد عملت أستاذا مساعدا، لمدة عام واحد، في جامعة سانت جونز، الولايات المتحدة الأمريكية، ثم عملت منذ سبتمبر ١٩٩٤، أستاذا في معهد بالمر لعلم المكتبات والمعلومات، في جامعة لونج أيلاند، الولايات المتحدة الأمريكية. ومع تقدمها في مسارها العلمي، تغير توجهها الموضوعي من علم المكتبات بوجه عام، إلى استرجاع المعلومات على وجه الخصوص. وقد ألفت خمسة كتب بالصينية، حول استرجاع المعلومات، وما يتصل به من موضوعات، كما أسهمت منذ عام ١٩٩٧ بسبعة كتب بين التأليف والتحرير. وتشمل اهتماماتها البحثية والتدريسية استخدام تقنيات المعلومات في المكتبات، ومناهج البحث، والاتصال العلمي. ولها أكثر من أربعين مقالة في موضوعات المكتبات الرقمية، والكتب والدوريات الإلكترونية، ومحركات البحث في العنكبوتية العالمية، وتطبيقات الجيل الثاني للعنكبوتية. ولها أيضا نشاطها العلمي والمهني واسع المدى، على المستويات المحلية والوطنية والدولية.

المترجم فى سطور:

حشمت محمد على قاسم

أستاذ علم المعلومات المتفرغ فى كلية الآداب - جامعة القاهرة. ولد فى عزبة قاسم، مركز بسيون، بمحافظة الغربية، في ١٦/ ٣/ ١٩٤٣. وقد تخرج في الوثائق والمكتبات، بمرتبة الشرف، فى كلية الآداب - جامعة القاهرة، دور يونيو ١٩٦٤، وحصل على الماجستير في التوثيق العلمى، فى الكلية نفسها عام ١٩٧١، وحصل على الدكتوراه في علم المعلومات، فى جامعة لندن عام ١٩٧٨. وقد تدرج في وظائف التدريس بكلية الآداب - جامعة القاهرة، منذ عام ١٩٦٤. أعير خبيرا للمكتبات، فى جامعة الإمارات العربية المتحدة، ومديرا لدار الكتب الوطنية في أبو ظبي، ثم أستاذا لجامعة الإمام محمد بن سعود الإسلامية فى الرياض. وقد عمل خبيرا ومستشارا في المكتبات وتنظيم المعلومات، في عدد من المؤسسات المصرية والعربية والدولية، كما شارك في كثير من المؤتمرات الوطنية والدولية، ورأس تحرير مجلة "دراسات عربية في المكتبات وعلم المعلومات". وله الكثير من الأعمال المنشورة، من الكتب المؤلفة، والكتب المترجمة، والبحوث والمقالات، ومراجعات الكتب. ويبدي في أعماله العلمية اهتماما خاصا بأصول علم المعلومات، والاتصال العلمى، وإدارة مرافق المعلومات، وخدمات المعلومات، وتقنيات المعلومات في المكتبات، والقياسات الوراقية والمعلوماتية بوجه عام. وقد حصل على جائزة معهد المعلومات العلمية ISI، على مجمل نشاطه العلمى، عام ٢٠٠٣.

التصحيح اللغوى: محمد أبو العز

الإشراف الفنـــى: حســن كــامل



تنظيم المعلومات أهم مقومات استثمار المعلومات، وهو مجال ديناميكي دائم التطور، نتيجة لارتباطه الوثيق بتقنيات المعلومات سريعة التغير. ويرصد الكتاب الوضع الراهن للمعرفة في مجال تنظيم المعلومات واسترجاعها، بجانبيه النظري والتطبيقي. ويبدأ بنظرة تاريخية موجزة، تربط الماضي بالحاضر، لاستشراف آفاق المستقبل. كما يتناول بأسلوب سلس بالغ الوضوح والإيجاز، خصائص المعلومات الرقمية، والأساليب والطرق والنظم والممارسات التي نشأت وتطورت في ظل الإنترنت. وفضلا عن اهتمامه بالجوانب التقنية، يبدي هذا الكتاب اهتماما واضحا بالمستفيد من المعلومات، من حيث احتياجاته وأنماط وسبل تفاعله مع نظم استرجاع المعلومات بكل فئاتها. وبالإضافة إلى الشمول في معالجة قضايا المجال، يوثق الكتاب ارتباط القارئ برصيد الإنتاج الفكري الثري، من خلال منظومة الاستشهادات المرجعية الموزعة على الفصول الاثني عشر. وكما يخاطب هذا الكتاب الدارسين والممارسين والباحثين في مجال تنظيم المعلومات واسترجاعها، فإنه يقدم للمستفيدين من المعلومات ما يسهم في توعيتهم ودعم ثقافتهم المعلوماتية.

تصميم الغلاف : جاسر ممدوح